Replicability analysis for genome-wide association studies

Ruth Heller and Daniel Yekutieli[1]

Abstract. The paramount importance of replicating associations is well recognized in the genome-wide associaton (GWA) research community, yet methods for assessing replicability of associations are scarce. Published GWA studies often combine separately the results of primary studies and of the follow-up studies. Informally, reporting the two separate meta-analyses, that of the primary studies and follow-up studies, gives a sense of the replicability of the results. We suggest a formal empirical Bayes approach for discovering whether results have been replicated across studies, in which we estimate the optimal rejection region for discovering replicated results. We demonstrate, using realistic simulations, that the average false discovery proportion of our method remains small. We apply our method to six type two diabetes (T2D) GWA studies. Out of 803 SNPs discovered to be associated with T2D using a typical meta-analysis, we discovered 219 SNPs with replicated associations with T2D. We recommend complementing a meta-analysis with a replicability analysis for GWA studies.

Keywords: Combined analysis; Empirical Bayes; False discovery rate; Meta-analysis; Replication; Reproducibility; Type 2 diabetes.

1   Introduction

The aim of a genome-wide association (GWA) study is to identify genetic variants that are associated with a given phenotype. An analysis that combines several GWA studies of the same phenotype may have increased power to discover the genetic variants that are associated with the phenotype. Such a meta-analysis combines all the data from all the studies to compute an overall p-value for each SNP. The overall p-values are used to identify the loci that are associated with the disease. A seminal example of combining data to identify association comes from the field of type 2 diabetes (T2D) GWAS. Voight et al. (2010) discover in a meta-analysis single nucleotide polymorphisms (SNPs) associated with T2D that were not discovered in single studies.

The paramount importance of replicating associations has been well-recognized in the GWAS literature (e.g. McCarthy et al., 2008; NCI-NHGRI, 2007). Kraft et al. (2009) note that for common variants, the anticipated effects are modest and very similar in magnitude to the subtle biases that may affect genetic association studies - most notably population stratification bias. For this reason, they argue that it is important to see the association in other studies conducted using a similar, but not identical, study base. Ioannidis and Khoury (2011) discuss multiple steps needed to validate "omics" findings, including "replication" which they define as the step to answer the question "Do many different data sets and their combination (meta-analysis) get consistent results?".

Meta-analysis of several GWA studies aims to discover the associations that are present in at least one study, not replicated associations. We define *replicability analysis* as an analysis with the aim to discover replicated associations, i.e. associations between SNP and phenotype that are present in more than one of the studies. Meta-analysis methods are not appropriate for discovering replicated associations. To see this, consider the scenario where for testing the null hypothesis that a SNP is independent of the phenotype, the p-value is extremely small in one study, but not small at all in the other studies. The meta-analysis will result in a small combined p-value, since there is evidence of association of this SNP with the phenotype, but there

is no evidence that this association is replicated. Therefore, a small p-value in a typical meta-analysis is evidence towards association of the SNP with the phenotype in at least one study, but it is not evidence that the association has been replicated in more than one study.

Many methods exist for meta-analysis, where follow-up studies simply serve to add power. See Hedges and Olkin (1985), Benjamini and Yekutieli (2005), Skol et al. (2006), and Zeggini et al. (2007), among others. However, only a handful of methods have been suggested so far for replicability analysis. Benjamini, Heller and Yekutieli (2009; hereafter, BHY09) suggest applying the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995), henceforth referred to as the BH procedure, on partial conjunction hypotheses p-values introduced in Benjamini and Heller (2008). Bogomolov and Heller (2012) focus on replicability analysis for two studies, and suggest an alternative false discovery rate (FDR) controlling procedure for this setting. Natarajan et al. (2012) suggest a list-intersection test to compare the top-ranked gene lists from multiple studies in order to discover a common significant set of genes. In this work, we suggest an empirical Bayes approach to replicability analysis. This approach may be viewed as an extension of the empirical Bayes approach of Efron (2008). We estimate the local Bayes FDRs under the various configurations of association status of SNP with phenotype across studies, and then sum up the relevant probabilities in order to estimate the Bayes FDR.

The motivating example for this work came from the field of T2D GWA studies, and therefore we discuss this work in the context of GWA studies. However, the proposed approach is a general approach for assessing replicability in several studies when each study examines the same hypotheses. Section 2 describes the motivating example, and defines formally our replicability analysis aim. In Section 3 we present the empirical Bayes method, and in Section 4 we apply the method to the motivating example. In Section 5 we use simulations to evaluate the performance of our method. We show that in realistic simulations, the average false discovery proportion (FDP) of our method remains small, while the power is much greater than the power of the method of BHY09. A similar observation was made in Sun and Wei (2011), where the advantage of using an empirical Bayes approach to testing sets of hypotheses over the method of Benjamini and Heller (2008) was illustrated by an application to time-course microarray data. We conclude with a brief summary in Section 6.

2 Motivating example and formulation of the replicability analysis aims

Voight et al. (2010) conducted a meta-analysis of eight T2D GWA studies comprising 8130 T2D cases and 38,987 controls of European descent. They combined the case-referent data from the Wellcome Trust Case Control Consortium (WTCCC), the Diabetes Genetics Initiative (DGI), the Finland-US Investigation of NIDDM genetics (FUSION) scans, deCode genetics (DECODE), the Diabetes gene Discovery Group, the Cooperative Health Research in the Region of Augsburg group, the Rotterdam study (ERGO), and the European Special Populatin Research Network (EUROSPAN). Based on a meta-analysis of these studies, Voight et al. (2010) selected few dozen SNPs for follow-up, and reported the SNPs that had a small p-value in the follow-up study, saying that these SNPs showed, in their words, "strong evidence for replication".

We received permission to use the p-values for the following six studies used for meta-analysis in Voight et al. (2010): EUROSPAN, DECODE, ERGO, DGI, FUSION, and WTCCC. For these six studies, our aim was to discover the SNPs that show strong evidence for replication of association with T2D within a formal statistical analysis framework. Replication of association can be defined in several ways: with or without regard to the direction of association; with at least u out of the six studies showing association, where $u \in \{2, \ldots, 6\}$ is fixed in advance. Since direction consistency is typically sought between the primary and follow-up studies in GWAS (e.g. Voight et al., 2010), our definition takes the directionality into account. For the six studies, we consider a SNP as having a replicated association if there is enough evidence to establish that the association of SNP with the phenotype is in the same direction in at least two studies.

In order to define the replicability aim formally, we use the following notation. Suppose there are n indepen-

dent studies, and in each study M SNPs are measured. For SNP j in study i, define H_{ij} as follows:

$$H_{ij} = \begin{cases} 1 & \text{if SNP } j \text{ is positively associated with the phenotype in study } i, \\ 0 & \text{if SNP } j \text{ is not associated with the phenotype in study } i, \\ -1 & \text{if SNP } j \text{ is negatively associated with the phenotype in study } i. \end{cases}$$

Let T_{ij} be the test statistic of SNP j in study i. Following Efron (2010), rather than computing the p-value, we transform the test statistic into a z-score $Z_{ij} = \Phi^{-1}(F_{i0}(T_{ij}))$, where F_{i0} is the cumulative distribution functions for T_{ij} when $H_{ij} = 0$ and Φ^{-1} is the inverse of the standard normal cumulative distribution function, respectively. The conditional density of Z_{ij} given H_{ij} is

$$f(z|H_{ij}) = \begin{cases} f_{i,1}(z) & \text{if } H_{ij} = 1, \\ f_0(z) & \text{if } H_{ij} = 0, \\ f_{i,-1}(z) & \text{if } H_{ij} = -1, \end{cases}$$

where $f_0(z)$ is the standard normal density.

Let $\mathcal{H} = \{\vec{h} = (h_1, \ldots, h_n) : h_i \in \{-1, 0, 1\}\}$ be the set of 3^n possible configurations of the vector of association status (of SNP with phenotype) in the n studies. We are interested in examining null hypotheses for the n studies that are defined by subsets of $\mathcal{H}$ denoted by $\mathcal{H}^0$. In particular, we shall examine the *no association null hypothesis* H_{NA}^0 that the SNP is not associated with the phenotype in any of the studies,

$$\mathcal{H}_{NA}^0 : \{(0, 0, \cdots, 0)\},$$

as well as the *no replicability null hypothesis* H_{NR}^0 that the SNP is positively and negatively associated with the phenotype in at most one study,

$$\mathcal{H}_{NR}^0 : \{\vec{h} : \sum_{i=1}^n I(h_i = -1) \le 1 \cap \sum_{i=1}^n I(h_i = 1) \le 1\},$$

where $I(\cdot)$ is the indicator function.

Our primary goal in this work is to discover as many SNPs as possible with false H_{NR}^0. This goal is distinct from the meta-analysis goal, of discovering as many SNPs as possible with false H_{NA}^0. For example, for $n = 2$ studies, $\mathcal{H}$ contains $3^2 = 9$ configurations,

$$\mathcal{H}_{NA}^0 = \{(0, 0)\},$$

$$\mathcal{H}_{NR}^0 = \{(0, 0), (1, 0), (0, 1), (-1, 0), (0, -1), (1, -1), (-1, 1)\},$$

and we aim to discover as many SNPs from the index set $\{j : \vec{H}_j \in \mathcal{H}/\mathcal{H}_{NR}^0\}$, where $\mathcal{H}/\mathcal{H}_{NR}^0 = \{(1, 1), (-1, -1)\}$. Had we defined replicability without taking directionality into account, the null hypothesis of interest would have been $\mathcal{H}^0 = \{(0, 0), (1, 0), (-1, 0), (0, 1), (0, -1)\}$, which aims to discover as many SNPs as possible from the index set $\{j : \vec{H}_j \in \{(1, 1), (-1, -1), (-1, 1), (1, -1)\}\}$. This aim could be pursued just as easily as the aim that follows from our definition of replicability, with the analysis method of the next Section 3, but we do not examine it here.

3 The empirical Bayes approach to replicability analysis

3.1 The empirical Bayes approach to multiple testing

The two group model provides a simple Bayesian framework for multiple testing, see e.g. Chapter 2 in Efron (2010). Each SNP in study i has marginal probability $\pi_0(i)$ of not being associated with the phenotype,

i.e. $\Pr(H_{ij} = 0) = \pi_0(i)$. Conditional on $H_{ij} = 0$, the SNP has a standard normal density, $f_0(z)$. Unconditionally, the continuous marginal (mixture) density is $f_i(z)$. For a subset $\mathcal{Z}$ of $\Re$, let $P_0(\mathcal{Z}) = \int_{\mathcal{Z}} f_0(z)dz$ and $P_i(\mathcal{Z}) = \int_{\mathcal{Z}} f_i(z)dz$.

Suppose we observe $z_{ij} \in \mathcal{Z}$ and wish to test $H_{ij} = 0$. A direct application of Bayes' theorem yields

$$Fdr_i(\mathcal{Z}) = Pr(H_{ij} = 0 | z_{ij} \in \mathcal{Z}) = \pi_0(i)P_0(\mathcal{Z})/P_i(\mathcal{Z}).$$

Adopting the terminology in Efron (2010), we call $Fdr_i(\mathcal{Z})$ the *Bayes FDR* for $\mathcal{Z}$: if we report $z_{ij} \in \mathcal{Z}$ as non-null, i.e. if we report $H_{ij} \neq 0$, then $Fdr_i(\mathcal{Z})$ is the chance that we have made a false discovery, i.e that $H_{ij} = 0$.

Theorem 1 of Storey (2003) shows that for the two group model for independent test statistics, $Fdr_i(\mathcal{Z})$ is closely connected to the FDR introduced in Benjamini and Hochberg (1995). Let $\mathbf{H}_i = (H_{i1}, \ldots, H_{iM})$, $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{iM})$, $Q(\mathcal{Z}, \mathbf{H}_i) = \sum_{j=1}^{M} I(z_{ij} \in \mathcal{Z}, H_{ij} = 0)/\max(R_i, 1)$, where $R_i = \sum_{j=1}^{M} I(z_{ij} \in \mathcal{Z})$ is the number of z-scores in the rejection region. The FDR is $FDR(\mathcal{Z}, \mathbf{H}_i) = E_{\mathbf{Z}_i | \mathbf{H}_i} Q(\mathcal{Z}, \mathbf{H}_i)$. Taking expectation over the random $\mathbf{H}_i$,

$$E_{\mathbf{H}_i}[FDR(\mathcal{Z}, \mathbf{H}_i)] = Pr(R_i > 0)E_{\mathbf{Z}_i, \mathbf{H}_i}[Q | R_i > 0] = Pr(R_i > 0)Fdr_i(\mathcal{Z}).$$

If $\mathcal{Z}$ is a single point z_0, then the *local Bayes FDR* is

$$fdr_i(z_0) = Pr(H_{ij} = 0 | z_{ij} = z_0) = \pi_0(i)f_0(z_0)/f_i(z_0).$$

$Fdr_i(\mathcal{Z})$ is the conditional expectation of $fdr_i(z)$ given $z \in \mathcal{Z}$ (Efron and Tibshirani, 2002),

$$Fdr_i(\mathcal{Z}) = E_{f_i}(fdr_i(z) | z \in \mathcal{Z}). \tag{1}$$

The Bayes false negative rate is $Fnr_i(\mathcal{Z}) = Pr(H_{ij} \neq 0 | z_{ij} \notin \mathcal{Z})$ (Efron, 2010). Similar to Storey (2007) and Sun and Cai (2007), we observe that among all possible rejection regions $\mathcal{Z}$ constrained to satisfy that $Fdr(\mathcal{Z}) \leq q$, the region with maximal probability, and with minimal Bayes false negative rate, will be of the form

$$\mathcal{Z}_{OR} = \{z : fdr_i(z) \leq t(q)\}. \tag{2}$$

The result is stated formally in the following proposition.

Proposition 3.1. *Assume the two group model holds for the z-scores in study i. Let $t(q)$ in expression (2) be such that $Fdr_i(\mathcal{Z}_{OR}) = q$. For any $\mathcal{Z}$ satisfying $Fdr_i(\mathcal{Z}) \leq q$,*

1. $P_i(\mathcal{Z}) \leq P_i(\mathcal{Z}_{OR})$.

2. $Fnr_i(\mathcal{Z}_{OR}) \leq Fnr_i(\mathcal{Z})$.

See the proof in Section 1 of the Supplementary Material.

In the two group model, $\pi_0(i)$ and f_i are needed in order to compute the local Bayes FDR. These quantities are estimated in the R package *locfdr*, available on CRAN. Poisson regression is used to estimate the marginal density of the z-scores, $\hat{f}_i$. The assumption that z-scores that fall in the range of the central 50% of the null distribution are null is used to estimate the fraction of null hypotheses: $\hat{\pi}_0(i) = \frac{|\{j : z_{ij} \in [\Phi^{-1}(0.25), \Phi^{-1}(0.75)]\}|}{M \times 0.5}$. Other estimation methods are suggested in Strimmer (2008), Muralidharan (2010), Storey and Tibshirani (2003), Benjamini et al. (2006) , and Jin and Cai (2007).

For a rejection region $\mathcal{Z}$, equation (1) shows that $Fdr_i(\mathcal{Z})$ may be estimated by

$$\widehat{Fdr}_i(\mathcal{Z}) = \frac{\sum_{j : z_{ij} \in \mathcal{Z}} \widehat{fdr}(z_{ij})}{|\{j : z_{ij} \in \mathcal{Z}\}|},$$

where $\widehat{fdr}_i(z_{ij}) = \hat{\pi}_0(i) f_0(z_{ij}) / \hat{f}_i(z_{ij})$ is the estimated local Bayes FDR of z-score z_{ij}, and $|\{j : z_{ij} \in \mathcal{Z}\}|$ is the number of z-scores in $\mathcal{Z}$. The estimated optimal rejection region is $\mathcal{Z} = \{z_{ij} : \widehat{fdr}_i(z_{ij}) \le t(q)\}$, where $t(q)$ is the largest threshold so that $\widehat{Fdr}_i(\mathcal{Z})$ is at most q.

3.2 Generalization of the two group model

Each SNP has probability $\pi(\vec{h})$ of having association configuration $\vec{h}$, i.e. $\Pr(\vec{H}_j = \vec{h}) = \pi(\vec{h})$. We assume the z-scores are independent across studies conditional on the association status $\vec{H}_j = \vec{h}$, so the vector of n z-scores $\vec{z}_j = (z_{1j}, \ldots, z_{nj})$ has density $f(\vec{z}_j|\vec{h}) = \prod_{i=1}^{n} f(z_{ij}|h_i)$. Note that $\pi_0(i)$ is equal to the sum of the probabilities $\pi(\vec{h})$ over all 3^{n-1} configurations $\vec{h} \in \mathcal{H}$ with $h_i = 0$.

Suppose we observe $\vec{z}_j$ for SNP j and wish to test $\vec{H}_j \in \mathcal{H}^0$. A direct application of Bayes' theorem yields the local Bayes FDR

$$fdr_{\mathcal{H}^0}(\vec{z}_j) = Pr(\vec{H}_j \in \mathcal{H}^0|\vec{z}_j) = \sum_{\vec{h} \in \mathcal{H}^0} \pi(\vec{h}) f(\vec{z}_j|\vec{h}) / f(\vec{z}_j), \tag{3}$$

where $f(\vec{z}_j) = \sum_{\vec{h} \in \mathcal{H}} \pi(\vec{h}) f(\vec{z}_j|\vec{h})$ is the mixture density. The local Bayes FDR for SNP j for null hypothesis H_{NA}^0 and H_{NR}^0, respectively, is

$$fdr_{\mathcal{H}_{NA}^0}(\vec{z}_j) = Pr(\vec{H}_j \in \mathcal{H}_{NA}^0|\vec{z}_j) \quad \text{and} \quad fdr_{\mathcal{H}_{NR}^0}(\vec{z}_j) = Pr(\vec{H}_j \in \mathcal{H}_{NR}^0|\vec{z}_j).$$

For a subset $\mathcal{Z}$ of $\Re^n$, if we report for $\vec{z}_j \in \mathcal{Z}$ that $\vec{H}_j \notin \mathcal{H}^0$, then the Bayes FDR is, as in equation (1),

$$Fdr_{\mathcal{H}^0}(\mathcal{Z}) = Pr(\vec{H}_j \in \mathcal{H}^0|\vec{z}_j \in \mathcal{Z}) = E_f(fdr_{\mathcal{H}^0}(\vec{z}_j)|\vec{z}_j \in \mathcal{Z}). \tag{4}$$

The optimal rejection region to discover SNPs that are non-null, i.e. $\vec{H}_j \notin \mathcal{H}^0$, follows from the same optimality argument of Proposition 3.1. The rejection region with maximal probability and minimal Bayes false negative rate among all possible rejection regions that are constrained to have a Bayes FDR of at most level q, is

$$\mathcal{Z}_{OR,\mathcal{H}^0} = \{\vec{z} : fdr_{\mathcal{H}^0}(\vec{z}) \le t(q)\}, \tag{5}$$

where $t(q)$ is such that $Fdr_{\mathcal{H}^0}(\mathcal{Z}_{OR,\mathcal{H}^0}) = q$. Section 2 of the Supplementary Material shows numerical examples that demonstrate the different optimal rejection regions for no replicability null hypotheses and for no association null hypotheses, as well as the loss in power that occurs when the rejection region is chosen sub-optimally based on p-values.

To test whether $\vec{H}_j \in \mathcal{H}^0$ on the n studies, we need to first estimate the local Bayes FDR for the observed z-scores, $\{\widehat{fdr}_{\mathcal{H}^0}(\vec{z}_k) : k = 1, \ldots, M\}$. We use these estimates to estimate the Bayes FDR (4) for every z-score $\vec{z}_j$ $(j = 1, \ldots, M)$:

$$\widehat{Fdr}_{\mathcal{H}^0}(\mathcal{Z}_j) = \frac{\sum_{k:\vec{z}_k \in \mathcal{Z}_j} \widehat{fdr}_{\mathcal{H}^0}(\vec{z}_k)}{|\{k : \vec{z}_k \in \mathcal{Z}_j\}|}, \tag{6}$$

where $\mathcal{Z}_j = \{\vec{z}_k : \widehat{fdr}_{\mathcal{H}_0}(\vec{z}_k) \le \widehat{fdr}_{\mathcal{H}_0}(\vec{z}_j), k = 1, \ldots, M\}$. Let $\hat{t}(q)$ be the largest estimated local Bayes FDR satisfying $\widehat{Fdr}(\mathcal{Z}_j) \le q$. Then, our estimate of the optimal rejection region (5) is $\{\vec{z}_k : \widehat{fdr}_{\mathcal{H}_0}(\vec{z}_k) \le \hat{t}(q), k = 1, \ldots, M\}$. We conclude that SNP k is non-null, i.e. $\vec{H}_k \notin \mathcal{H}_0$, if $\widehat{fdr}_{\mathcal{H}_0}(\vec{z}_k) \le \hat{t}(q)$, or equivalently, if $\widehat{Fdr}(\mathcal{Z}_k) \le q$.

To compute $f(\vec{z}_j)$ it is necessary to specify the conditional distributions for the three states of nature for association for each SNP in each study: $H_{ij} \in \{-1, 0, 1\}$. This is a key difference from the analysis

5

of single studies, where estimation of the marginal density of the z-scores does not require estimation of the conditional distributions. In Section 3 of the Supplementary Material we demonstrate the necessity of estimating the conditional distributions for the states $H_{ij} = -1$ and $H_{ij} = 1$ in order to get a good estimate of $f(\vec{z}_j)$ at the tails, for $\vec{H}_j$ with dependent components.

Next, we show how to estimate $\pi(\vec{h})$ and the conditional z-score densities that are necessary for estimating the local Bayes FDR.

3.3 Estimating $\pi(\vec{h})$ and the conditional z-score densities

The likelihood for the z-scores for SNP j is

$$L(\vec{\pi}; \vec{z}_j, f) = \Pr(\vec{z}_j | \vec{\pi}) = \sum_{\vec{h} \in \mathcal{H}} f(\vec{z}_j | \vec{h}) \pi(\vec{h}), \tag{7}$$

where $\vec{\pi} = \{\pi(\vec{h}) : \vec{h} \in \mathcal{H}, \sum_{\vec{h} \in \mathcal{H}} \pi(\vec{h}) = 1\}$ is the set of $3^n - 1$ probabilities of the multi-group model we want to estimate.

The full likelihood requires both the joint distribution of $(\vec{H}_1 \cdots \vec{H}_M)$ and, for each study i $(i = 1, \ldots, n)$, the joint distribution of $(Z_{i1}, \ldots, Z_{iM})$ given $(H_{i1}, \ldots, H_{iM})$. Since the joint distribution is unknown, we consider instead the composite likelihood, which is the product of the marginal likelihoods for the M SNPs,

$$L^{CL}(\vec{\pi}; \vec{z}, f) = \Pi_{j=1}^{M} L(\vec{\pi}; \vec{z}_j, f).$$

Although the composite likelihood is different than the full likelihood, in large problems with local dependency the maximum likelihood estimates of the composite likelihood and the full likelihood are very similar (Cox and Reid, 2004). For GWAS the assumption of local dependency seems reasonable, since the dependency across SNPs diminishes as the distance between the SNPs increases. In Section 5 we verified that the composite likelihood was indeed appropriate using simulated data with GWA dependency.

Assuming that the probabilities in $\vec{\pi}$ were known, the composite likelihood could be computed if the probability distributions of z_{ij} given $H_{ij} \in \{-1, 0, 1\}$, $i = 1, \ldots, n$, were known, since $f(\vec{z}_j | \vec{H}_j) = \prod_{i=1}^{n} f_{i,H_{ij}}(z_{ij})$. Conditional on $H_{ij} = 0$, the density of z_{ij}, denoted by $f_0(\cdot)$, is indeed known to be standard normal (in Section 6 we discuss what can be done when $f_0(\cdot)$ is unknown). Mixture model density estimation methods can be used to estimate $f_{i,1}$ and $f_{i,-1}$ (McLachlan, 2000). First, the methods discussed in Section 3.1 can be used to estimate the marginal density of the z-scores for each study, f_i, and the fraction of SNPs with no association with the phenotype, $\pi_0(i)$. Denoting the estimates by $\hat{f}_i$ and $\hat{\pi}_0(i)$, the bimodal alternative density is $\hat{f}_{i,A}(z) = \frac{\hat{f}_i(z) - \hat{\pi}_0(i) f_0(z)}{1 - \hat{\pi}_0(i)}$. Next, the expectation maximization (EM) algorithm, detailed in Section 4 of the Supplementary Material, is used to find $\vec{\pi}$ that maximizes the composite likelihood.

4 Replicability analysis of T2D GWA studies

Our first step in this analysis is to estimate the fraction of null hypotheses for each of the six studies, using the *locfdr* package. In two of the studies, the estimated fraction of null hypotheses is 1. Since a stable estimate of the conditional distribution under the alternative could not be extracted for these two studies, we excluded them from the empirical Bayes analysis. Studies DECODE, DGI, FUSION, and WTCCC had estimated fractions of null hypotheses of 0.89, 0.98, 0.98, and 0.96, respectively. Figure 1 of the Supplementary Material shows the histogram of z-scores, as well as the estimated conditional densities, for each of the six studies, as outputted from the *locfdr* package.

Binning of z-scores In the *locfdr* package, the z-scores are binned before the densities are estimated. Binning is practical in our application since in the estimation of the local Bayes FDRs for several studies, estimated conditional densities are multiplied. The accuracy of multiplied estimates may be far less stable without binning. Therefore, we first divide the z-scores $\{z_{ij} : j = 1, \ldots, M\}$ into B bins of equal width. For this application, we tried both $B = 50$ and $B = 120$ and received similar results. Let $x_{i,1} \cdots x_{i,B}$ be the centers of these bins. We assign each z-score z_{ij} into the bin that it is in, denoted by $\tilde{z}_{ij} \in \{1, \ldots, B\}$. For SNP j, the probability of the vector of n binned z-scores $\tilde{z}_j = (\tilde{z}_{1j}, \ldots, \tilde{z}_{nj})$ given configuration $\vec{H}_j$ is $\tilde{f}(\tilde{z}_j | \vec{H}_j) = \prod_{i=1}^{n} \tilde{f}_{i,H_{ij}}(\tilde{z}_{ij})$, where $\tilde{f}_{i,H_{ij}}(b) = \frac{f_{H_{ij}}(x_{i,b})}{\sum_{l=1}^{B} f_{H_{ij}}(x_{i,l})}$. For $H_{ij} = 0$, $f_0(x_{i,b})$ is the standard normal density at point $x_{i,b}$. For $H_{ij} \in \{-1, 1\}$,

$$f_{i,1}(x_{i,b}) = \begin{cases} 0 & \text{if } x_{i,b} \leq 0, \\ \hat{f}_A(x_{i,b}) & \text{if } x_{i,b} > 0. \end{cases} \quad \text{and} \quad f_{i,-1}(x_{i,b}) = \begin{cases} 0 & \text{if } x_{i,b} \geq 0, \\ \hat{f}_A(x_{i,b}) & \text{if } x_{i,b} < 0. \end{cases}$$

The EM algorithm was used to find $\vec{\pi}$ that maximizes the composite likelihood on the binned z-scores, $\prod_{j=1}^{M} \sum_{\vec{h} \in \mathcal{H}} \tilde{f}(\tilde{z}_j | \vec{h}) \pi(\vec{h})$.

For $n = 4$ studies, the sets $\mathcal{H}$ and $\mathcal{H}_{NR}^0$ contain, respectively 81 and 21 configurations, and $\mathcal{H}_{NA}^0$ contains only the configuration $(0, 0, 0, 0)$. The empirical Bayes analysis at level $q = 0.05$ discovered 803 SNPs associated with T2D and 219 SNPs with replicated association with T2D. A list of the 219 SNPs with replicated associations discovered by the empirical Bayes analysis, sorted by positions on the chromosome, is given in Section 5 of the Supplementary Material. SNPs with replicated association included 16 distinct genes. We extracted the SNP with smallest estimated local Bayes FDR among all SNPs within each of these 16 genes, as well as among all SNPs in non-coding areas. In Table 1 we list these 17 SNPs, along with the estimated Bayes FDR for replicability analysis (column 5) and for the analysis to discover association in (column 6). As expected, the estimated Bayes FDR is larger for replicability analysis than for an analysis to discover associations, and the ranking for replicability is different than for discovering associations. For example, the empirical Bayes analysis for KIF11 ranks it 7th for evidence of replicability but 5th for evidence of association; KCNJ11 is ranked 5th for evidence of replicability but 8th for evidence of association. The SNP which has by far the strongest evidence of association, and replicated association, is in TCF7L2. This association has been well established in previous studies (Voight et al., 2010). The very small estimated Bayes FDRs for this SNP are a result of compounding the strong evidence against the null from four studies.

As a comparison procedure, we considered the replicability analysis suggested in BHY09, which was to apply the BH procedure on the M no replicability null hypotheses p-values, computed as suggested in Benjamini and Heller (2008). We applied the analysis suggested in BHY09 on the $n = 4$ studies with estimated fraction of null hypotheses below one, as well as on all the $n = 6$ studies available. Briefly, the recipe for computing p-values for the no replicability null hypotheses was as follows. First, for every subset of $n - 1$ studies, a meta-analysis p-value was computed. Then, the p-value for the no replicability null hypothesis was set to be the maximum of the n meta-analysis p-values. Since we considered in this work a concordant version of replicability, where the association was considered replicated only if it was present in at least two studies in the same direction, the p-value was taken to be twice the smaller of the left- and right-sided combined p-values using the method of Fisher, as suggested in Owen (2009).

The replicability analysis of BHY09 at level $q = 0.05$ based on the four studies, discovered 447 SNPs associated with T2D and 83 SNPs with replicated association with T2D, and based on the six studies discovered 466 SNPs associated with T2D and 113 SNPs with replicated association with T2D. Table 1 shows the adjusted p-values based on all six available studies in columns seven and eight, respectively. While the meta-analysis of BHY09 indicates that there is evidence of association in almost all these regions, evidence of replicated association is inferred only for five regions.

The empirical Bayes approach provides for each SNP a measure of belief in each possible configuration $\vec{h}$ conditional on its vector of z-scores. For example, the vector of z-scores for SNP $rs7903146$ in gene TCF7L2 was $\vec{z} = (-8.8, -4.5, -4.4, -7.5)$ in studies DECODE, DGI, FUSION, and WTCCC, respectively. The estimated posterior probability was 0.98 that the configuration was $\vec{h} = (-1, -1, -1, -1)$, conditional

on the binned z-score vector. The vector of z-scores for SNP $rs10923931$ in gene NOTCH2 was $\vec{z} = (-3.4, -4.9, -0.12, -2.8)$ with estimated posterior probability 0.92 for configuration $\vec{h} = (-1, -1, 0, -1)$. Table 2 shows the estimated posterior probability distributions for these two SNPs.

Table 1: For the SNPs with strongest evidence towards replicability in 17 distinct regions discovered by the empirical Bayes replicability analysis: the estimated Bayes FDR for replicability and for association (column 5-6); the adjusted p-values from the analysis of BHY09 for replicability and for association (column 7-8).

	chr	pos	gene	Empirical Bayes Fdr		BHY09 adjusted p-values	
				Replicability	Association	Replicability	Association
rs7903146	10	114758349	TCF7L2	2.40e-11	4.61e-22	0.00e+00	0.00e+00
rs10440833	6	20688121	CDKAL1	1.60e-05	8.06e-08	9.06e-09	0.00e+00
rs5015480	10	94465559	non-coding	1.10e-03	7.74e-05	8.78e-04	1.12e-07
rs4402960	3	185511687	IGF2BP2	3.14e-03	6.87e-04	0.0205	3.51e-05
rs5215	11	17408630	KCNJ11	8.91e-03	4.50e-03	1.00e+00	0.0236
rs757110	11	17418477	ABCC8	9.98e-03	6.16e-03	1.00e+00	0.0267
rs4933734	10	94414567	KIF11	0.0111	2.96e-04	1.00e+00	1.55e-05
rs10923931	1	120517959	NOTCH2	0.0134	2.70e-03	1.00e+00	3.45e-04
rs11187033	10	94262359	IDE	0.0189	2.07e-03	0.0186	7.07e-06
rs319602	5	134222164	TXNDC15	0.0202	7.07e-03	1.00e+00	0.0364
rs849134	7	28196222	JAZF1	0.0210	7.80e-03	9.84e-01	1.16e-03
rs6883047	5	134272055	PCBD2	0.0235	8.55e-03	1.00e+00	0.0471
rs10832778	11	17394073	B7H6	0.0282	0.0164	1.00e+00	1.53e-01
rs13070993	3	12217797	SYN2	0.0370	0.0235	1.00e+00	0.0369
rs10433537	3	12198485	TIMP4	0.0360	0.0233	1.00e+00	0.0386
rs10113282	8	96038252	C8orf38	0.0387	0.0102	1.00e+00	0.0408
rs1554522	17	25913172	KSR1	0.0436	0.0145	1.00e+00	2.13e-01

Table 2: The estimated posterior probabilities for different configurations $\vec{h}$, conditional on the binned z-score of $\vec{z}$, for two example z-scores: $rs7903146$ in gene TCF7L2 (column 2), and $rs10923931$ in gene NOTCH2 (column 3).

$\vec{h}$	$\vec{z} = (-8.8, -4.5, -4.4, -7.5)$	$\vec{z} = (-3.4, -4.9, -0.12, -2.8)$
(-1, -1 , -1 , -1)	0.980	0.000
(-1 , -1 , 0 , -1)	0.012	0.924
(-1 , -1 , 0 , 0)	0.000	0.047
(-1 , 0 , -1 , -1)	0.008	0.000
(-1 , 0 , 0 , -1)	0.000	0.004
(0, -1, 0, -1)	0.000	0.024
(0 , -1, 0, 0)	0.000	0.001

5 Simulation studies

If all parameters were known, the optimal rejection region could be calculated. In Section 2 of the Supplementary Material , we present two simple examples that demonstrate the difference between the optimal rejection region for a replicability analysis and that for an analysis to discover associations, and show that the optimal region can be much larger than that based on p-values. Since the optimal rejection region has to be estimated in practice, we examine here the empirical Bayes approach, that estimates the optimal rejection region for inference. Specifically, the goal of the simulations was twofold. First, to investigate the effect of

the number of SNPs M, and the dependence across SNPs, on the empirical Bayes procedure. Second, to compare the empirical Bayes procedure to the replicability analysis of BHY09 at the same level q. In the empirical Bayes analysis, the z-scores were first binned, using $B = 50$ bins, and SNPs were considered discovered if the estimated Bayes FDR in equation (6) was below $q = 0.05$. In addition to the empirical Bayes procedure that estimates $\vec{\pi}$ via the EM algorithm, we also considered the oracle Bayes procedure that knows the association status H_{ij} of each SNP. The oracle Bayes procedure estimates the conditional probabilities of the binned z-scores in each study by the relative frequency of each bin conditional on the association status, and uses the true vector $\vec{\pi}$ for computing the local Bayes FDRs.

5.1 Independence within each study

We considered $n = 3$ studies, with 2000 cases and 2000 referents and $M \in \{10^3, 10^4, 10^5\}$ SNPs in each study. Although there were $3^n = 27$ possible configurations of the vector of associations status, our data generation process had positive probability only for the 15 configurations that do not have a positive and negative association for the same SNP: configuration $(0, 0, 0)$ for 90% of the SNPs; the six configurations with exactly one true association, i.e. $\vec{H}_j$ s.t. $\sum_{i=1}^{3} |H_{ij}| = 1$, each for 1% of the SNPs; the eight configurations with at least two true associations in the same direction, i.e. $\vec{H}_j$ s.t. $|\sum_{i=1}^{3} H_{ij}| \geq 2$, each for 0.5% of the SNPs. Following Wakefield (2007), we simulated data for every SNP independently with disease risk, p_{ij}, given by the logistic regression model $logit(p_{ij}) = \alpha + u\theta_{ij}$, where $u = 0, 0.5$, and 1 corresponds to 0,1 and 2 copies of the mutant allele, respectively. We sample θ_{ij} given H_{ij} as follows:

$$\theta_{ij}|H_{ij} \sim \begin{cases} U(0.25, 0.5) & \text{if } H_{ij} = 1, \\ 0 & \text{if } H_{ij} = 0, \\ U(-0.5, -0.25) & \text{if } H_{ij} = -1. \end{cases}$$

where $U(a, b)$ denotes the uniform distribution between a and b. Moreover, the minor allele frequency (MAF) for each SNP j in study i, was sampled from $U(0.05, 0.50)$, and we set $\alpha = -6$, so $e^\alpha = 0.0025$ was the prior odds of a disease due to a SNP with $u = 0$.

Results The simulation results were based on 50 repetitions for $M = 10^5$, and on 100 repetitions for $M = 10^4$ and $M = 1000$. Figure 2 in the Supplementary Material shows the FDP in an analysis to discover associations and in a replicability analysis. The variation in FDP decreases with M, and is very small for $M = 10^5$. Table 3 presents the average FDP, and number of rejections, R. Although the average FDP of the empirical Bayes analysis was below 0.05 for $M \geq 10^4$, the average FDP when $M = 1000$ was 0.071, with a standard error (SE) of 0.006. The empirical Bayes analysis makes only few more discoveries than the the analysis of BHY09 when the aim is to discover associations, but three-fold more discoveries when the aim is to discover replicated associations. For example, for $M = 10^5$ SNPs the empirical Bayes analysis discovers on average 2040 SNPs with replicated associations, while the analysis of BHY09 discovers only an average of 684 SNPs. A comparison of columns 6 and 8 shows that the oracle Bayes analysis produces only few more discoveries than the empirical Bayes analysis, suggesting that the loss of power in the estimation of the parameters is small.

Remark 5.1. *Table 3 shows that the average FDP for the analysis of BHY09 when the aim is to discover associations was lower than $\pi(0, 0, 0) \times 0.05 = 0.045$. For example, for $M = 10^5$ the average FDP was 0.039. This is due to the discreteness of the distribution of the p-values, that were computed from contingency tables. Indeed, when the sample size was tripled, the p-values from true no association null hypotheses were closer to uniform and therefore the average FDP was closer to the nominal level (not shown). However, the over-conservativeness of the replicability analysis remained severe when the sample size was tripled.*

Table 3: The average FDP and number of rejections R, in an empirical Bayes analysis (columns 3 and 6), in the analysis of BHY09 (columns 4 and 7), and in an oracle Bayes analysis (columns 5 and 8), for different values of M=number of hypotheses.

Analysis type	M	FDP ($SE \times 1000$)			R (SE)		
		Empirical Bayes	BHY09	Oracle Bayes	Empirical Bayes	BHY09	Oracle Bayes
Replicability	10^5	0.049 (1)	0.001 (0)	0.050 (1)	2040.6 (6.3)	684.1 (3.4)	2091.6 (4.8)
	10^4	0.049 (2)	0.000 (0)	0.049 (1)	203.6 (1.4)	68 (0.9)	211.2 (1.1)
	10^3	0.071 (6)	0.000 (0)	0.044 (4)	20.5 (0.4)	7.1 (0.3)	22.7 (0.3)
Association	10^5	0.046 (0)	0.039 (0)	0.050 (0)	5911.3 (8.7)	5495.8 (7.8)	6047.0 (9.3)
	10^4	0.047 (1)	0.038 (1)	0.050 (1)	591.3 (1.7)	549.7 (1.8)	610.6 (1.8)
	10^3	0.051 (2)	0.040 (3)	0.045 (2)	58.7 (0.6)	54.9 (0.6)	66.6 (0.5)

5.2   GWA dependency within each study

We simulated three GWA studies from the simulator HAPGEN2 (Su et al., 2011). The three studies where generated from three samples of the HapMap project (The International HapMap Consortium, 2003): a sample of 87 individuals with African ancestry in Southwest USA (ASW), a sample of 165 Utah residents with Northern and Western European ancestry (CEU), and a sample of 109 Chinese in Metropolitan Denver, Colorado (CHD). We limited ourselves to chromosomes 1-4, that contained $M = 415,154$ SNPs. In these populations, the number of causal SNPs was 26 for ASW, 22 for CEU and 27 for CHD. Since the effects are typically small for GWA studies, we consider for each population four sub-populations, and within each sub-population about 1/4 of the causal SNPs had an increased multiplicative relative risk of 1.5. Overall, there were 48 different causal SNPs in the four chromosomes, out of which 22 SNPs were causal in more than one population. Specifically, the three populations had five causal SNPs in common, and in addition, the number of causal SNPs in common in exactly two of the three populations was: four for ASW and CEU, seven for ASW and CHD, and six for CEU and CHD. Each study contained 8000 cases and 8000 referents from each population. The simulator HAPGEN2 uses an estimate of the fine-scale recombination rate map to simulate haplotypes conditional on the reference haplotype data from the HapMap project. The simulator assumes a hidden Markov model and treats the recombination rates and mutation rates as transition and emission probabilities, respectively. The resulting simulated data has the same linkage disequilibrium (LD) patterns as each reference data from the HapMap project.

Due to LD, the number of SNPs associated with the phenotype in every study was larger than the number of causal SNPs. Since it is not known from the data generation process which SNPs are associated with the phenotype in each study, then for a non-causal SNP j we do not know whether $H^0 \in \{H^0_{NA}, H^0_{NR}\}$ is false, since non-causal SNPs may have false H^0 due to LD patterns in the different populations. Since a major goal in the simulations was to assess whether the FDP is inflated, it was necessary to establish a ground truth. We wanted to estimate a conservative ground truth that with very high probability estimates a SNP as having a true H^0 if indeed it is from H^0, at the possible expense of estimating a SNP as having a true H^0 even if H^0 was false. The estimation of the ground truth was as follows. The simulation studies were repeated 20 times, resulting in 20 p-values per population for every SNP. The 20 p-values were first combined with Fisher's combining method, and the analysis of BHY09 was applied to the combined p-values from the three populations, to form for each SNP a combined p-value for $H^0 \in \{H^0_{NA}, H^0_{NR}\}$ that is based on 20 studies per population. H^0 was considered to be false for a SNP if the p-value for testing H^0 was below the severe Bonferroni threshold for FWER control at level 0.05. The resulting ground truth contains 2126 SNPs associated with the phenotype, i.e. with false H^0_{NA}, and 695 SNPs with replicated association with the phenotype, i.e. with false H^0_{NR}. The ground truth based on 20 repetitions was very similar to a ground truth that was established based on only 19 of the 20 repetitions, and therefore for an analysis of one repetition, the resulting FDP using the ground truth based on 20 repetitions was very similar to the FDP using the ground truth that results from the 19 repetitions excluding the repetition being analyzed.

Results Table 4 shows the analysis results for the 20 repetitions of the three studies. Although the average number of rejections was only slightly larger with the empirical Bayes analysis than with the analysis of BHY09 for testing associations, it was more than 20 times larger when testing for replicated associations. The average FDP for the empirical Bayes analysis was slightly above the nominal level of 0.05, possibly because either "ground truth" was too conservative ("false rejections" are not really "false") or the empirical Bayes analysis is indeed slightly anti-conservative for the type of dependency that occurs in GWA studies. Nevertheless, this simulation demonstrates the large gain in using an empirical Bayes analysis over the analysis of BHY09 for discovering replicated associations. This large gain comes at a small risk, slightly inflated FDP.

Table 4: The average FDP, and number of rejections R, in an empirical Bayes analysis (columns 2 and 4), and in the analysis of BHY09 (columns 3 and 5), for the simulated data with GWA dependency within each study.

| | FDP ($SE \times 1000$) | | R (SE) | |
Analysis type	Empirical Bayes	BHY09	Empirical Bayes	BHY09
Replicability	0.065 (9)	0.000 (0)	154.1 (8.5)	6.4 (1.2)
Association	0.072 (9)	0.053 (5)	274.9 (12.4)	242.7 (10.4)

6   Summary

In our analysis, we assumed for each study that if the null hypothesis was true for a SNP, the p-values was uniformly distributed, i.e. the z-score had a standard normal density. Efron (2008) lists several reasons why the empirical null may be preferred over the theoretical null distribution of the z-scores. The R package *locfdr* fits the empirical null by truncated maximum likelihood or by fitting a quadratic to $\log f_i$ near the center. If in doubt about the theoretical null, the theoretical null may be replaced with the empirical null in the empirical Bayes analysis. In our analysis we estimated the conditional density of Z_{ij} given $H_{ij} \in \{-1, 0, 1\}$ in order to discover replicated positive and negative associations. In future work we intend to examine a more general parametrization of the associations.

The accuracy of the empirical Bayes analysis relies on the ability to estimate well the unknown parameters. We demonstrated in simulations that the variability of the FDP decreased as the number of hypotheses increased. In a simulation of realistic GWA studies we demonstrated that the empirical Bayes analysis produced inferences with a small FDP, despite the dependency among the p-values within each study. A full Bayesian approach to the problem of GWA studies replicability is not possible, since we do not know the true likelihood. To estimate the probabilities of each of the 3^n configurations of null and non-null hypotheses, we used the product of the marginal SNP likelihoods. In applications were the exact likelihood is known, it is possible to use a full Bayesian approach, so that the suggested framework for replicability analysis can be extended to account for the uncertainty of the Bayes FDR estimates.

From a comparison of an empirical Bayes analysis with the analysis of BHY09, we see that they may give similar inferences when the analysis is aimed at discovering associations. However, for replicability the empirical Bayes analysis discovers many more replicated associations than the analysis of BHY09. In our analysis of the T2D studies, we removed the two studies with an estimated fraction of null hypotheses of one from the empirical Bayes analysis, since the alternative distribution could not be reliably estimated for these two studies using the R package *locfdr*. However, these studies are useful, as indicated by the fact that the analysis of BHY09 detected more associations using all 6 studies than using only the 4 studies with an estimated fraction of null hypotheses below one. How to best incorporate these two studies into the empirical Bayes analysis is an open question.

Acknowledgements

References

Benjamini, Y. and Heller, R. (2008). Screening for partial conjunction hypotheses. *Biometrics*, 64:1215–1222.

Benjamini, Y., Heller, R., and Yekutieli, D. (2009). Selective inference in complex research. *Philosophical Transactions of the Royal Society A*, 267:1–17.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B Met.*, 57 (1):289–300.

Benjamini, Y., Krieger, M., and Yekutieli, D. (2006). Adaptive linear step-up false discovery rate controlling procedures. *Biometrika*, 93 (3):491–507.

Benjamini, Y. and Yekutieli, D. (2005). Quantitative trait loci analysis using the false discovery rate. *Genetics*, 171:783–790.

Bogomolov, M. and Heller, R. (2012). Discovering findings that replicate from a primary study of high dimension to a follow-up study. *arXiv:1207.0187v1*.

Cox, D. and Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, 91 (3):729–737.

Efron, B. (2008). Microarrays, empirical bayes and the two-groups model. *Statistical Science*, 23:1–22.

Efron, B. (2010). *Large-Scale Inference*. Cambridge, United Kingdom.

Efron, B. and Tibshirani, R. (2002). Empirical bayes methods and false discovery rates for microarrays. *Genetic Epidemiology*, 23:70–86.

Hedges, L. and Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Academic Press, London.

Ioannidis, J. and Khoury, M. (2011). Improving Validation Practices in "Omics" Research. *Science*, 334 :1230–1232.

Jin, J. and Cai, T. (2007). Estimating the null and the proportion of nonnull effects in large-scale multiple comparisons. *Journal of the American Statistical Association*, 102 (478):495–506.

Kraft, P., Zeggini, E., and Ioannidis, J. (2009). Replication in genome-wide association studies. *Statistical science*, 24 (4):561–573.

McCarthy, M., Abecasis, G., Cardon, L., Goldstein, D., Little, J., Ioannidis, J., and Hirschhorn, J. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature reviews Genetics*, pages 356–369.

McLachlan, G., an Peel, D. (2010). *Finite Mixture Models*. John Wiley & Sons, USA.

Muralidharan, O. (2010). An empirical Bayes mixture method for effect size and false discovery rate estimation. *The Annals of Applied Statistics*, 4(1), pages 422–438.

Natarajan, L., Pu, M., and Messer, K. (2012). Statistical tests for the intersection of independent lists of genes: sensitivity, FDR, and type I error control. *The Annals of Applied Statistics*, 6 (2):521–541.

NCI-NHGRI (2007). Replicating genotype-phenotype associations. *Nature*, 447(7):655–660.

Owen, A. (2009). Karl pearson's meta-analysis revisited. *The annals of statistics*, 37 (6B):3867–3892.

Skol, A., Scott, L., Abecasis, G., and Boehnke, M. (2006). Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nature Genetics*, 38:209–213.

Storey, J. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B*, 64 (3):79–498.

Storey, J. (2003). The Positive False Discovery Rate: A Bayesian Interpretation and the q-Value. *The Annals of Statistics*, 31(6):2013–2035.

Storey, J. (2007). The optimal discovery procedure: A new approach to simultaneous significance testing. *Journal of the Royal Statistical Society, Series B*, 69:347–368.

Storey, J. and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, 100 (16):9440–9445.

Strimmer, K. (2011). A unified approach to false discovery rate estimation. *BMC Bioinformatics*, 9 (303).

Su, Z., Marchini, J., and Donnelly, P. (2011). Hapgen2: simulation of multiple disease snps. *Bioinformatics*, 27 (16):2304–2305.

Sun, W. and Cai, T. (2007). Oracle and adaptive compound decision rules for false discovery rate control. *Journal of the American Statistical Association*, 102 (479):901–912.

Sun, W. and Wei, Z. (2011). Multiple testing for pattern identification, with application to microarray time-course experiments. *Journal of the American Statistical Association*, 106 (493):73–88.

The International HapMap Consortium (2003). The International Hapmap Project. *Nature*, 426:789–796.

Voight et al. (2010). Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature Genetics*, 42:579–589.

Wakefield, J. (2007). A bayesian measure of the probability of false discovery in genetic epidemiology studies. *The American Journal of Human Genetics*, 81:208–227.

Zeggini, E., Weedon, M., Lindgren, C., Frayling, T., Elliott, K., Lango, H., Timpson, N. Perry, J., and Rayner, N. (2007). Replication of genome-wide association signals in uk samples reveals risk loci for type 2 diabetes. *Science*, 316:1336–1341.

Supplementary Material for Replicability analysis for Genome-wide Association studies

S7 Proof of Proposition 3.1

Proof. Since the result is for a single study i, for notational convenience we omit the subscript i in the following proof. Let $\psi_{OR}(z)$ be the indicator of whether $z \in \mathcal{Z}_{OR}$, and let $\psi(z)$ be the indicator of whether $z \in \mathcal{Z}$ for another rejection region that satisfies $Fdr(\mathcal{Z}) \leq q$. Straightforward calculus shows for every z

$$\psi(z)(1 - fdr(z)/t(q)) \leq \psi_{OR}(z)(1 - fdr(z)/t(q)) \tag{8}$$

Taking expectations on both sides of equation (8),

$$\int [\psi(z)(1 - fdr(z)/t(q))]f(z)dz \leq \int [\psi_{OR}(z)(1 - fdr(z)/t(q))]f(z)dz,$$

we receive the following expression:

$$P(\mathcal{Z})(1 - Fdr(\mathcal{Z})/t(q)) \leq P(\mathcal{Z}_{OR})(1 - Fdr(\mathcal{Z}_{OR})/t(q)) \tag{9}$$

Since $Fdr(\mathcal{Z}_{OR})$ is the expectation of $fdr(z)$ for $fdr(z) \leq t(q)$, and $q < t(q)$, it follows that $Fdr(\mathcal{Z}_{OR}) < t(q)$. Moreover, since $Fdr(\mathcal{Z}) \leq Fdr(\mathcal{Z}_{OR})$, it follows that $(1 - Fdr(\mathcal{Z})/t(q)) \geq (1 - Fdr(\mathcal{Z}_{OR})/t(q)) > 0$. Therefore, the right hand side of expression (9) is smaller than $P(\mathcal{Z}_{OR})(1 - Fdr(\mathcal{Z})/t(q))$ and item 1 follows.

In order to prove item 2, let $tdr(z) = 1 - fdr(z)$ be the true discovery rate. Straightforward calculus shows for every z

$$[1 - \psi(z)][1 - tdr(z)/(1 - t(q))] \leq [1 - \psi_{OR}(z)][1 - tdr(z)/(1 - t(q))] \tag{10}$$

Taking expectations on both sides of equation (10),

$$\int [1 - \psi(z)][1 - tdr(z)/(1 - t(q))]f(z)dz \leq \int [1 - \psi_{OR}(z)][1 - tdr(z)/(1 - t(q))]f(z)dz,$$

we receive the following expression:

$$[1 - P(\mathcal{Z})][1 - Fnr(\mathcal{Z})/(1 - t(q))] \leq [1 - P(\mathcal{Z}_{OR})][1 - Fnr(\mathcal{Z}_{OR})/(1 - t(q))] \tag{11}$$

Since $fdr(z) > t(q)$ for $z \notin \mathcal{Z}_{OR}$, it follows that $1 - tdr(z)/(1 - t(q)) > 0$ for $z \notin \mathcal{Z}_{OR}$, and therefore that $1 - Fnr(\mathcal{Z}_{OR})/(1 - t(q)) > 0$. Combining this observations with the fact from item 1 that $1 - P(\mathcal{Z}) \geq 1 - P(\mathcal{Z}_{OR})$, the RHS of equation (11) can be bounded above by $[1 - P(\mathcal{Z})][1 - Fnr(\mathcal{Z}_{OR})/(1 - t(q))]$. It thus follows that $1 - Fnr(\mathcal{Z})/(1 - t(q)) \leq 1 - Fnr(\mathcal{Z}_{OR})/(1 - t(q))$, proving item 2. ∎

S8 Testing normal means

In this section we give simple examples that demonstrate that the rejection region for replicability analysis is very different than for an analysis to discover associations, and also that the optimal rejection regions may be far larger than a rejection region based on p-values. In this section only, for simplicity, we assume that each hypothesis has only two states: the null state with zero expectation, and the non-null state with positive expectation.

S8.1 Comparison of Bayes FDR for optimal and p-value based Bayesian analysis

Example S8.1. *For $n = 2$ studies, suppose the marginal z-score density in the two studies is $N(0, 1)$ under the no-association null hypothesis, and under the alternative positive association hypotheses the z-score density is $N(\mu_1, 1)$ in the first study and $N(\mu_2, 1)$ in the second study. Thus the joint z-score density is*

$$
\begin{aligned}
f(z_1, z_2) &= \pi(0,0)\phi(z_1)\phi(z_2) + \pi(0,1)\phi(z_1)\phi(z_2 - \mu_2) \\
&+ \pi(1,0)\phi(z_1 - \mu_1)\phi(z_2) + \pi(1,1)\phi(z_1 - \mu_1)\phi(z_2 - \mu_2),
\end{aligned}
$$

for $\phi(z)$ the standard Normal density. For $\vec{h} \in \{(0,0), (0,1), (1,0), (1,1)\}$, the conditional probability that $\vec{H} = (h_1, h_2)$ given (z_1, z_2) is

$$
\Pr(\vec{H} = \vec{h}| (z_1, z_1)) = \frac{\pi(\vec{h})\phi(z_1 - I(h_1 = 1)\mu_1)\phi(z_2 - I(h_2 = 1)\mu_2)}{f(z_1, z_2)}.
$$

The local Bayes fdr for testing H_{NA}^0 and H_{NR}^0, respectively, is $fdr_{NA}(z_1, z_2) = \Pr(\vec{H} = (0,0)| (z_1, z_1))$ and

$$
fdr_{NR}(z_1, z_2) = \Pr(\vec{H} = (0,0)| (z_1, z_2)) + \Pr(\vec{H} = (0,1)| (z_1, z_2)) + \Pr(\vec{H} = (1,0)| (z_1, z_2)).
$$

We compared the optimal rejection region and the rejection region based on p-values for the Bayesian analysis. The p-values for H_{NA}^0 and H_{NR}^0 were, respectively, the p-values of the Fisher combined (right-sided) p-values,and the maximum of the two studies p-values. Specifically, for $\vec{Z} = (Z_1, Z_2)$, let $P_1 = 1 - \Phi(Z_1)$ and $P_2 = 1 - \Phi(Z_2)$. The p-value for testing no-association was $P^{NA} = 1 - F_{\chi_4^2}(-2(log(P_1) + log(P_2)))$, and the p-value for testing no-replication was $P^{NR} = \max(P_1, P_2)$.

For $\mu_1 = \mu_2 = 3$, let $\pi(0,0) = 0.80, \pi(0,1) = \pi(1,0) = 0.08$, and $\pi(1,1) = 0.04$. Figure S1 shows the rejection region boundaries for the optimal rejection region (solid) and the p-values based rejection region (dashed) for testing H_{NA}^0 (top) and H_{NR}^0 (middle). Clearly, the rejection regions are much larger for detecting associations than for detecting replicability.

For $\pi(0,0) = 0.88, \pi(0,1) = 0.12$, Figure S1 (bottom) shows the rejection regions when testing H_{NA}^0. The hypothesis H_{NR}^0 is not tested, since the local Bayes FDR of no replicability is one, and there does not exist a region with Bayes FDR at most $q < 1$. The difference between the optimal rejection region and the rejection region based on p-values is much larger in this configuration than in the previous configuration. Specifically, the optimal rejection region is only determined by the z-score of the second study, Z_2.

Table S5 shows the probability of the rejection regions for the no association and no replicability null hypotheses. The probabilities of the rejection regions to discover replicability are much smaller than for discovering associations. Moreover, the probabilities of the optimal rejection regions are larger than for the p-value based region, and the differences between the probabilities of the regions are larger for configuration $\pi(0,0) = 0.88, \pi(0,1) = 0.12$ than for $\pi(0,0) = 0.80, \pi(0,1) = \pi(1,0) = 0.08$, and $\pi(1,1) = 0.04$.

The following example illustrates the large loss of power due to a non-optimal choice of rejection region that can occur when more than two studies are available.

Example S8.2. *For $n = 6$ studies, let $\pi((0,0,0,0,0,0)) = 0.90$ and $\pi((0,0,0,0,0,1)) = 0.10$. Thus the first five z-scores $Z_1 \cdots Z_5$ are $N(0,1)$. The sixth z-score Z_6 is $N(0,1)$ with probability 0.9 and $N(3,1)$ with probability 0.1. Similar to the setting $(\mu_1, \mu_2) = (0,3)$ in Example S8.1, the p-value based rejection region for testing H_{NA}^0 is very different than the optimal rejection region, which is only based on Z_6. For a Bayes FDR of $q = 0.05$, the probability of the optimal rejection region was 0.066, and the probability of the p-value based rejection region was 0.012.*
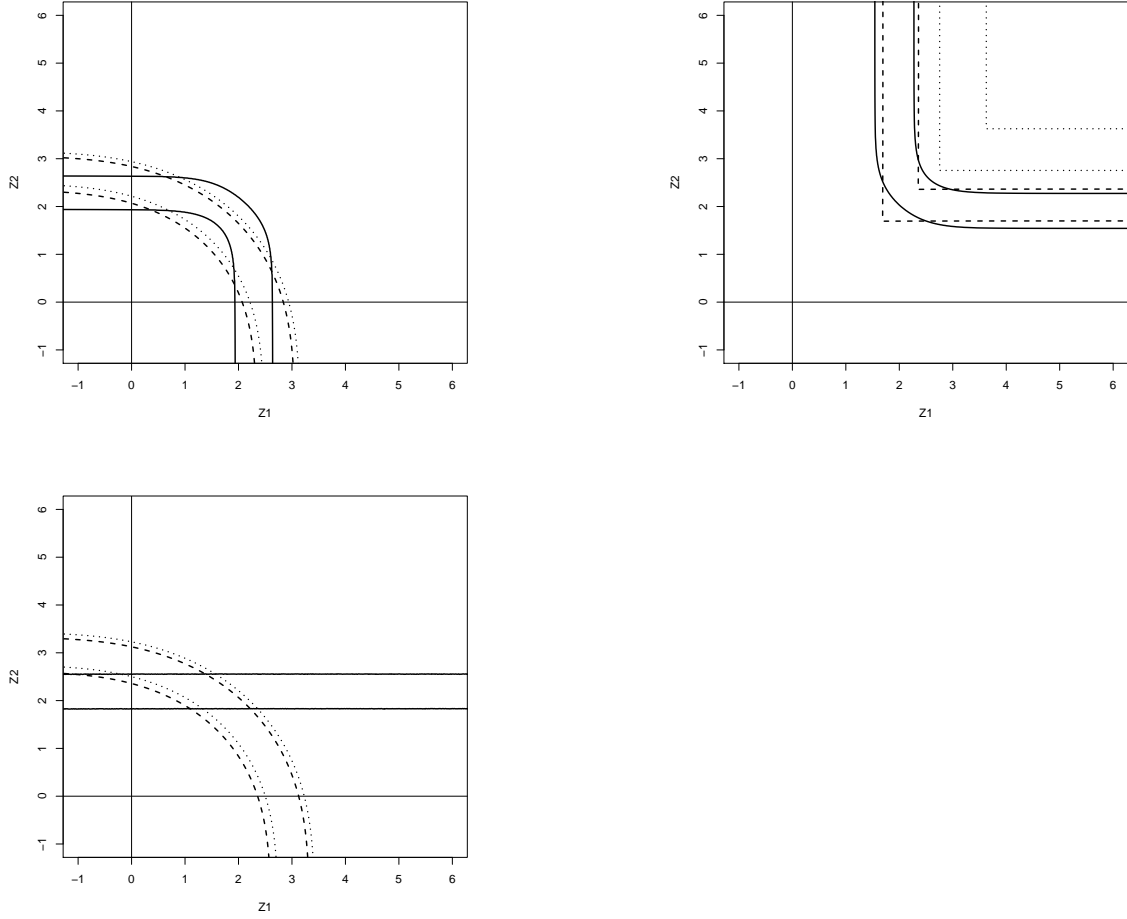
Figure S1: Optimal (solid curves), p-value based (dashed curves) rejection regions boundaries for bayes FDR levels $q \in \{0.20, 0.05\}$, as well as the rejection region for the analysis of Benjamini et al. (2009) (dotted curves) for FDR levels $q \in \{0.20, 0.05\}$, in configuration $\pi((0,0)) = 0.80, \pi((0,1)) = \pi((1,0)) = 0.08,,$ and $\pi((1,1)) = 0.04$ of the test of H_{NA}^0 (top) and H_{NR}^0 (middle), and in configuration $\pi((0,0)) = 0.88, \pi((0,1)) = 0.12$ for the test of H_{NA}^0 (bottom). The further the boundary is from $(0,0)$ the smaller the value of q.

Table S5: The probability of the optimal and of the p-value based rejection regions, for various Bayes FDR levels q and two configurations of $\pi = (\pi(0,0), \pi(0,1), \pi(1,0), \pi(1,1))$.

Null	$\pi(0,0), \pi(0,1), \pi(1,0), \pi(1,1)$	Rejection region	$q = 0.05$	$q = 0.20$
H_{NR}^0	$(0.80, 0.08, 0.08, 0.04)$	$\mathcal{Z}_{OR}$	0.0234	0.0417
		p-value: Bayes	0.0230	0.0410
		BHY09	0.0028	0.0145
H_{NA}^0	$(0.80, 0.08, 0.08, 0.04)$	$\mathcal{Z}_{OR}$	0.1498	0.2230
		p-value: Bayes	0.1417	0.2182
		BHY09	0.1334	0.2007
H_{NA}^0	$(0.88, 0.12, 0.00, 0.00)$	$\mathcal{Z}_{OR}$	0.0855	0.1355
		p-value: Bayes	0.0621	0.1178
		BHY09	0.0563	0.1050

S8.2   Comparison of Bayes FDR for p-value based Bayesian analysis and for the BH procedure

In Example S8.1, the dotted curve in Figure S1 shows the rejection region using the BH procedure, as suggested in BHY09. While the rejection region is only slightly smaller than that of the p-value based Bayesian rejection region for testing for no association (top and bottom), it is much smaller for testing for no replicability (middle figure). We shall explain why these differences arise.

In the two group model, when the rejection region is based on the tails of the z-scores $\mathcal{Z} = \{z : z \le t(q)\}$ which are equivalent to one-sided p-values, there is a strong connection between empirical Bayes estimation of the Bayes FDR and the frequentist BH procedure for FDR control, as noted by Efron and Tibshirani (2002) and Storey (2002). If the jth p-value in study i is $p_{ij} = \Phi(z_{ij})$, then the BH rule rejects all hypotheses with z-scores that satisfy the following inequality:

$$\widehat{Fdr}_{i,BH}(z_{i(j)}) = \max_{l \ge j} \Phi(z_{i(l)})/(l/M) \le q, \tag{12}$$

where $z_{i(j)}$ is the jth largest z-score in study i. Since j/M is the empirical distribution of $\mathcal{Z}$ for the rejection region $\mathcal{Z}_j = \{z : z \le z_{i(j)}\}$, then if we set $\pi_0(i)$ conservatively to be one, the BH procedure coincides with the procedure that chooses the largest $\mathcal{Z}_j$ so that the estimated $Fdr_i(\mathcal{Z}_j)$ is at most q. Specifics follow. The rejection region of the BH procedure is $\mathcal{Z}_{\hat{p}_{i,BH}} = \{\vec{z} : P_{ij} \le \hat{p}_{BH}\}$, where $\hat{p}_{i,BH} = \sup\{p : \widehat{Fdr}_{i,BH}(p) \le q\}$ and

$$\widehat{Fdr}_{i,BH}(p) = \frac{p}{|\{j : p_{ij} \le p\}|/M}. \tag{13}$$

The Bayes FDR of $\mathcal{Z}_p = \{z_{ij} : P_{ij} \le p\}$ is

$$Fdr_i(\mathcal{Z}_p) = \Pr(H_{ij} = 0 | P_{ij} \in \mathcal{Z}_p) = \frac{\pi_0(i)\Pr(P_{ij} \le p | H_{ij} = 0)}{\Pr(P_{ij} \in \mathcal{Z}_p)} \tag{14}$$

Comparing (13) with (14), as the denominator of (13) is the empirical distribution of the event in the denominator of (14), if P_{ij} is $U[0,1]$ under the null hypothesis, the Fdr estimator in (13) is too large by a factor of $1/\pi_0(i)$. If P_{ij} is stochastically greater than $U[0,1]$, the Fdr estimator in (13) may be greatly over-conservative.

Similarly, for the null hypothesis $\mathcal{H}_{NA}^0$, the conservative factor is $1/\pi(\vec{0})$, since the rejection region of the BH procedure is $\mathcal{Z}_{\hat{p}_{BH}^{NA}} = \{\vec{z}_j : p_j^{NA} \le \hat{p}_{BH}^{NA}\}$, where $\hat{p}_{BH}^{NA} = \sup\{p : \widehat{Fdr}_{BH}^{NA}(p) \le q\}$ and

$$\widehat{Fdr}_{BH}^{NA}(p) = \frac{p}{|\{\vec{z}_j : p_j^{NA} \le p\}|/M}, \tag{15}$$

and the Bayes FDR of $\mathcal{Z}_p = \{\vec{z} : p_j^{NA} \le p\}$ is

$$Fdr^{NA}(\mathcal{Z}_p) = \Pr(\mathcal{H}_{NA}^0 | \vec{P}_j \in \mathcal{Z}_p) = \frac{\pi(\vec{0})\Pr(P^{NA} \le p | \vec{H} = \vec{0})}{\Pr(\vec{z}_j \in \mathcal{Z}_p)} \tag{16}$$

However, $\mathcal{H}_{NR}^0$ is a composite null hypothesis and therefore the conservativeness of the BH procedure is far greater. The rejection region of the BH procedure is $\mathcal{Z}_{\hat{p}_{BH}^{NR}} = \{\vec{z}_j : p_j^{NR} \le \hat{p}_{BH}^{NR}\}$, where $\hat{p}_{BH}^{NR} = \sup\{p : \widehat{Fdr}_{BH}^{NR}(p) \le q\}$ and

$$\widehat{Fdr}_{BH}^{NR}(p) = \frac{p}{|\{\vec{z}_j : p_j^{NR} \le p\}|/M}. \tag{17}$$

The Bayes FDR of $\mathcal{Z}_p = \{\vec{z} : p_j^{NR} \le p\}$ is

$$Fdr^{NR}(\mathcal{Z}_p) = \Pr(\mathcal{H}_{NR}^0 | \vec{z}_j \in \mathcal{Z}_p) = \frac{\sum_{\vec{h} \in \mathcal{H}_{NR}^0} \pi(\vec{h})\Pr(P^{NR} \le p | \vec{H} = \vec{h})}{\Pr(\vec{z}_j \in \mathcal{Z}_p)} \tag{18}$$

17
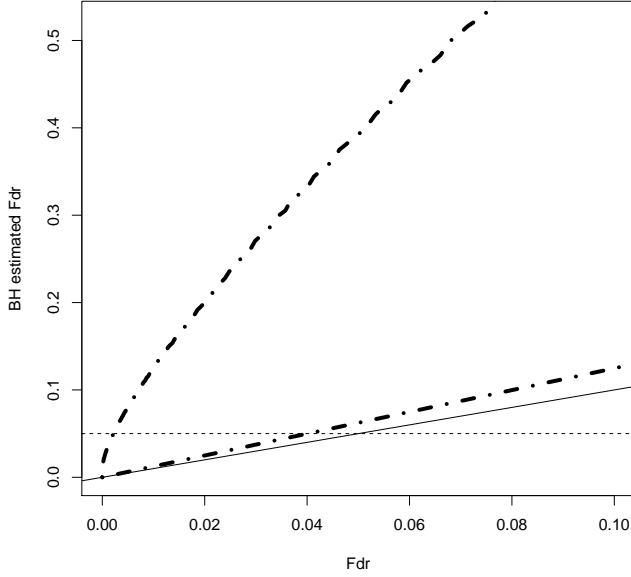
Figure S2: The Bayes FDR value q (solid line), $\widehat{Fdr}_{BH}$ for the test of H^0_{NA} (bottom dash-dot line) and for the test of H^0_{NR} (top dash-dot line). The horizontal dashed line is at level 0.05, and it intersects the bottom and top dash-dot lines at $(Fdr, \widehat{Fdr}_{BH})$ values (0.05,0.05) and (0.0022, 0.05), respectively.

Comparing (17) with (18), as the denominator of (17) is the empirical distribution of the event in the denominator of (18), the conservatism of the BH procedure follows from the differences in the numerators of these two expressions. The BH procedure is conservative since P^{NR}_j is stochastically greater than $U[0,1]$ for $\vec{H} \in \mathcal{H}_0$, especially for $\vec{H} = \vec{0}$. Therefore, the numerator in (17) is much larger than the numerator in (18) when $\pi(\vec{0})$ is large.

For Example S8.1, Table S5 shows the probability of the BH rejection regions for $(\mu_1, \mu_2) = (3,3)$. Figure S2 shows $\widehat{Fdr}^{NA}_{BH}(p)$ versus $Fdr^{NA}(\mathcal{Z}_p)$ and $\widehat{Fdr}^{NR}_{BH}(p)$ versus $Fdr^{NR}(\mathcal{Z}_p)$. For testing H^0_{NA}, $\widehat{Fdr}_{BH}$ was overly conservative by a factor of 1.25. Therefore the rejection region with $\widehat{Fdr}_{BH} = 0.05$ actually had $Fdr = 0.04$, and a rejection probability of 0.133, while the rejection probability was 0.1417 for the p-value based rejection region with $Fdr = 0.05$. For testing H^0_{NR}, the rejection region with $\widehat{Fdr}_{BH} = 0.05$ actually had $Fdr = 0.0022$, and the rejection probability was only 0.0028. For comparison, the rejection probability was 0.0230 for the p-value based rejection region with $Fdr = 0.05$.

S9   Computation of $f(\vec{z})$

The locfdr package estimates $f(z_{ij} | \ H_{ij} = 0)$ and $\Pr(H_{ij} = 0)$ in addition to $f(z_{ij})$, and then derives $f(z_{ij} | \ H_{ij} \neq 0)$ through the relation

$$f(z_{ij}) = f(z_{ij} | \ H_{ij} = 0) \cdot \Pr(H_{ij} = 0) + f(z_{ij} | \ H_{ij} \neq 0) \cdot \{1 - \Pr(H_{ij} = 0)\}.$$

In replicability analysis, that considers

$$
\begin{aligned}
\Pr(\vec{H}_j = \vec{h}|\, \vec{z}_j) &= \frac{f(\vec{z}_j,\, \vec{H}_j = \vec{h})}{f(\vec{z}_j)} \\
&= \frac{\{\Pi_{i=1}^n f(z_{ij}|\, H_{ij} = h_i)\} \cdot \Pr(\vec{H}_j = \vec{h})}{f(\vec{z}_j)}
\end{aligned}
\tag{19}
$$

for $\vec{h} \neq \vec{0}$, it is also necessary to specify $f(z_{ij}|\, H_{ij} = -1)$ and $f(z_{ij}|\, H_{ij} = 1)$.

If H_{ij} are independent then the components of $\vec{z}_j$ are also independent and thus the locfdr estimates of the marginal z-score densities are sufficient for computing

$$
\Pr(\vec{H}_j = \vec{0}|\, \vec{z}_j) = \frac{\Pi_{i=1}^n \{f(z_{ij}|\, H_{ij} = 0) \cdot \Pr(H_{ij} = 0)\}}{\Pi_{i=1}^n f(z_{ij})}.
$$

However, if the components of $\vec{H}_j$ are dependent then specifying $f(z_{ij}|\, H_{ij} = -1)$ and $f(z_{ij}|\, H_{ij} = 1)$ is necessary for computing $\Pr(\vec{H}_j = \vec{0}|\, \vec{z}_j)$, as illustrated in the example below.

Example S9.1. *Assume that either $\vec{H}_j = \vec{0}$ or $\vec{H}_j \in \mathcal{H}^1$, for $\mathcal{H}^1 = \{-1, 1\}^n$, and let $\Pr(\vec{H}_j = \vec{0}) = \pi_0$. Therefore, $f(\vec{z}_j, \vec{H}_j = \vec{0}) = \{\Pi_{i=1}^n f(z_{ij}|\, H_{ij} = 0)\} \cdot \pi_0$. Since $f(\vec{z}_j) = f(\vec{z}_j, \vec{H}_j = \vec{0}) + f(\vec{z}_j, \vec{H}_j \in \mathcal{H}^1)$, to compute $\Pr(\vec{H}_j = \vec{0}|\, \vec{z}_j)$ we need $f(\vec{z}_j, \vec{H}_j \in \mathcal{H}^1)$. In general,*

$$
\begin{aligned}
f(\vec{z}_j, \vec{H}_j \in \mathcal{H}^1) &= \sum_{\vec{h} \in \mathcal{H}^1} f(\vec{z}_j, \vec{H}_j = \vec{h}) \\
&= \sum_{\vec{h} \in \mathcal{H}^1} f(\vec{z}_j|\, \vec{H}_j = \vec{h}) \cdot \Pr(\vec{H}_j = \vec{h}) \\
&= \sum_{\vec{h} \in \mathcal{H}^1} \{\Pi_{i=1}^n f(z_{ij}|\, H_{ij} = h_i)\} \cdot \Pr(\vec{H}_j = \vec{h}).
\end{aligned}
\tag{20}
$$

If the components of $\vec{H}_j$ were independent conditional on $\vec{H}_j \in \mathcal{H}^1$, then

$$
\begin{aligned}
f(\vec{z}_j, \vec{H}_j \in \mathcal{H}^1) &= f(\vec{z}_j|\, \vec{H}_j \in \mathcal{H}^1) \cdot \Pr(\vec{H}_j \in \mathcal{H}^1) \\
&= \{\Pi_{i=1}^n f(z_{ij}|\, H_{ij} \neq 0)\} \cdot (1 - \pi_0).
\end{aligned}
\tag{21}
$$

Note that to compute (20) it is necessary to estimate $f(z_{ij}|\, H_{ij} = -1)$ and $f(z_{ij}|\, H_{ij} = 1)$ and that if H_{ij} are independent conditional on $\vec{H}_j \in \mathcal{H}^1$ then expressions (20) and (21) are the same, but for large n and highly dependent H_{ij} they may be very different. To see this, we further assume $\Pr(\vec{H}_j = (1, \ldots, 1)) = (1 - \pi_0)/2$ and $\Pr(\vec{H}_j = (-1, \ldots, -1)) = (1 - \pi_0)/2$, and consider $\vec{z} = (z_{1j} \cdots z_{nj})$ with $0 < z_{ij}$ for which $f(z_{ij}|\, H_{ij} = -1) << f(z_{ij}|\, H_{ij} = 1)$ and $f(z_{ij}|\, H_{ij} = 0) << f(z_{ij}|\, H_{ij} = 1)$. Since $f(z_{ij}|\, H_{ij} = -1) << f(z_{ij}|\, H_{ij} = 1)$, expression (20) can be approximated as follows:

$$
\begin{aligned}
&\sum_{\vec{h} \in \mathcal{H}^1} \{\Pi_{i=1}^n f(z_{ij}|\, H_{ij} = h_i)\} \cdot \Pr(\vec{H}_j = \vec{h}) \\
&\approx \{\Pi_{i=1}^n f(z_{ij}|\, H_{ij} = 1)\} \cdot (1 - \pi_0)/2.
\end{aligned}
\tag{22}
$$

Furthermore since $\Pr(H_{ij} = 1|\, H_{ij} \neq 0) = 1/2$ and

$$
\begin{aligned}
f(z_{ij}|\, H_{ij} \neq 0) &= f(z_{ij}|\, H_{ij} = -1) \cdot \Pr(H_{ij} = -1|\, H_{ij} \neq 0) \\
&\quad + f(z_{ij}|\, H_{ij} = 1) \cdot \Pr(H_{ij} = 1|\, H_{ij} \neq 0),
\end{aligned}
$$

then $f(z_{ij}|\, H_{ij} = 1)/f(z_{ij}|\, H_{ij} \neq 0) \approx 2$. Thus expression (22) is $2^{(n-1)}$ larger than expression (21). Since $f(z_{ij}|\, H_{ij} = 0) << f(z_{ij}|\, H_{ij} = 1)$, it follows that $f(\vec{z}_j) \approx f(\vec{z}_j, \vec{H}_j \in \mathcal{H}^1)$. As the denominator of $\Pr(\vec{H}_j = \vec{0}|\, \vec{z}_j)$ is approximately $f(\vec{z}_j, \vec{H}_j \in \mathcal{H}^1)$ then in this case $\Pr(\vec{H}_j = \vec{0}|\, \vec{z}_j)$ is $2^{(n-1)}$ smaller then it would have been if H_{ij} were independent conditional on $\vec{H}_j \in \mathcal{H}^1$.

S10 The EM algorithm

The observed data are z-scores $\vec{z}_1, \ldots, \vec{z}_M$ and the missing values are $\vec{H}_1, \ldots, \vec{H}_M$. The complete likelihood for SNP j is

$$L_c(\vec{\pi}; \vec{z}_j, \vec{f}, \vec{H}_j) = f(\vec{z}_j | \vec{H}_j)\pi(\vec{H}_j).$$

The composite complete likelihood for all the SNPs is

$$\Pi_{j=1}^M L_c(\vec{\pi}; \vec{z}_j, \hat{f}, \vec{H}_j) = \Pi_{j=1}^M f(\vec{z}_j | \vec{H}_j)\pi(\vec{H}_j).$$

E step In the E step we calculate the expected value of the log composite likelihood function, with respect to the conditional distribution of H given $\vec{z}$ under the current estimate of the parameters, $\vec{\pi}^{(t)}$:

$$
\begin{aligned}
Q(\vec{\pi} | \, \vec{\pi}^{(t)}) &= E_{H|\vec{z},\vec{\pi}^{(t)}}[log\{\Pi_{j=1}^M f(\vec{z}_j|\vec{H}_j) \cdot \pi(\vec{H}_j)\}] \\
&= \sum_{j=1}^M \sum_{\vec{h} \in \mathcal{H}} \Pr(\vec{H}_j = \vec{h}|\vec{z}_j, \vec{\pi}^{(t)})[\log f(\vec{z}_j|\vec{H}_j = \vec{h}) + \log\{\pi(\vec{h})\}] \\
&= \sum_{j=1}^M \sum_{\vec{h} \in \mathcal{H}} \Pr(\vec{H}_j = \vec{h}|\vec{z}_j, \vec{\pi}^{(t)})logf(\vec{z}_j|\vec{H}_j = \vec{h}) \\
&\quad + \sum_{j=1}^M \sum_{\vec{h} \in \mathcal{H}} \Pr(\vec{H}_j = \vec{h}|\vec{z}_j, \vec{\pi}^{(t)}) \log\{\pi(\vec{h})\}
\end{aligned}
\tag{23}
$$

where

$$\Pr(\vec{H}_j = \vec{h}|\vec{z}_j, \vec{\pi}^{(t)}) = \frac{f(\vec{z}_j|\vec{h})\pi^{(t)}(\vec{h})}{\sum_{\vec{h'} \in \mathcal{H}} f(\vec{z}_j|\vec{h'},)\pi^{(t)}(\vec{h'})}$$

M step Find $\vec{\pi}^{(t+1)}$ that maximizes $Q(\vec{\pi} | \, \vec{\pi}^{(t)})$. Since the second sum in equation (23) has the same form as the log-likelihood for the multinomial distribution, it follows that

$$\pi^{(t+1)}(\vec{h}) = \frac{\sum_{j=1}^M \Pr(\vec{H}_j = \vec{h}|\vec{z}_j, \vec{\pi}^{(t)})}{\sum_{\vec{h'} \in \mathcal{H}} \sum_{j=1}^M \Pr(\vec{H}_j = \vec{h'}|\vec{z}_j, \vec{\pi}^{(t)})}.$$

The updated parameters are $\vec{\pi}^{(t+1)} = \{\pi^{(t+1)}(\vec{h}) : \vec{h} \in \mathcal{H}\}$.

starting value $\pi^{(0)}$ As starting values, we recommend using values constrained to satisfy $\hat{\pi}_0(i) = \sum_{\{\vec{h} \in \mathcal{H}\} \cap \{h_i=0\}} \pi^{(0)}(\vec{h})$. Such a starting position will provide a good initial estimate of the non-null densities in the E step. Specifically, given estimates $\hat{\pi}_0(i), i = 1, \ldots, n$, we suggest as starting values

$$\pi^{(0)}(\vec{h}) = \Pi_{i=1}^n \hat{\pi}_{h_i}(i),$$

where $\hat{\pi}_1(i) = \hat{\pi}_{-1}(i) = (1 - \hat{\pi}_0(i))/2$.

Updating of $f(\vec{z}_j|\vec{h})$ After the EM converged to a new estimate, the estimated fraction of null hypotheses in each study can be extracted: $\hat{\pi}_0^{(T)}(i) = \sum_{\{\vec{h}\in\mathcal{H}\}\cap\{h_i=0\}} \pi^{(T)}(\vec{h})$, where T is the number of steps till convergence of the EM. A modified estimate of $f_{i,1}$ and $f_{i,-1}$ can then be computed using the new estimates $\hat{\pi}_0^{(T)}(i), i = 1, \ldots, n$, if these estimates are different than the starting values $\hat{\pi}_0(i), i = 1, \ldots, n$. These modified estimated can now be used to recompute $f(\vec{z}_j|\vec{h})$. Next, the EM can be repeated with the new estimated conditional densities. This iterative process should end when the new estimates of $\hat{\pi}_0^{(T)}(i), i = 1, \ldots, n$, are almost the same as the starting values of the EM.

S11   Replicability analysis of T2D GWA studies

Figure S3 shows the empirical z-scores, as well as the estimated conditional densities, for each of the six studies, as outputted from the locfdr package.

The table below gives the list of the 219 SNPs with replicated associations, as discovered by the empirical Bayes analysis, sorted by positions on the chromosome. The positions were found by NCBI build GRCh37.p5 reference assembly, and they were mapped to nearby genes by dbSNP (`http://www.ncbi.nlm.nih.gov/projects/SNP/dbSNP.cgi?list=rslist`). The table shows the estimated Bayes FDR for replicability analysis as well as for the analysis to discover association, and the adjusted p-values from the corresponding analysis of BHY09 based on all six available studies.

	chr	pos	gene	Empirical Bayes Fdr		BHY09 adjusted p-values	
				Repl.	Assoc.	Repl.	Assoc.
rs10923931	1	120517959	NOTCH2	1.34e-02	2.70e-03	1.00e+00	3.45e-04
rs6442307	3	12143355	SYN2	4.43e-02	2.74e-02	1.00e+00	4.89e-02
rs11715886	3	12147236	SYN2	4.40e-02	2.73e-02	1.00e+00	4.89e-02
rs4488811	3	12182028	SYN2	3.74e-02	2.36e-02	1.00e+00	4.21e-02
rs11721223	3	12185160	SYN2	3.67e-02	2.35e-02	1.00e+00	4.11e-02
rs11708978	3	12188495	SYN2	3.64e-02	2.34e-02	1.00e+00	4.06e-02
rs6792867	3	12189900	SYN2	3.77e-02	2.37e-02	1.00e+00	4.02e-02
rs7629805	3	12192394	SYN2	4.79e-02	3.15e-02	1.00e+00	5.10e-02
rs10433537	3	12198485	SYN2,TIMP4	3.60e-02	2.33e-02	1.00e+00	3.86e-02
rs13070993	3	12217797	SYN2	3.70e-02	2.35e-02	1.00e+00	3.69e-02
rs11720578	3	12267084	non-coding	4.33e-02	2.68e-02	1.00e+00	4.81e-02
rs13071168	3	12275447	non-coding	1.39e-02	1.24e-02	1.00e+00	1.53e-02
rs11709119	3	12276493	non-coding	4.14e-02	2.97e-02	1.00e+00	1.56e-02
rs17036101	3	12277845	non-coding	1.47e-02	1.26e-02	1.00e+00	1.56e-02
rs1562040	3	12285405	non-coding	1.43e-02	1.25e-02	1.00e+00	1.78e-02
rs17036130	3	12288288	non-coding	1.51e-02	1.27e-02	1.00e+00	1.73e-02
rs13081389	3	12289800	non-coding	4.17e-02	2.98e-02	1.00e+00	1.73e-02
rs1596417	3	12290898	non-coding	4.20e-02	2.99e-02	1.00e+00	1.75e-02
rs13089415	3	12301360	non-coding	2.27e-02	1.62e-02	1.00e+00	1.88e-02
rs6771792	3	12301472	non-coding	2.31e-02	1.63e-02	1.00e+00	1.82e-02
rs4376068	3	185497635	IGF2BP2	7.82e-03	2.16e-03	1.07e-01	1.88e-04
rs6801848	3	185499057	IGF2BP2	1.19e-02	3.81e-03	3.10e-01	9.60e-04
rs4481184	3	185505787	IGF2BP2	3.92e-03	1.12e-03	3.00e-02	5.22e-05
rs11705729	3	185507299	IGF2BP2	3.29e-03	8.83e-04	2.44e-02	4.23e-05
rs11929397	3	185510190	IGF2BP2	5.88e-03	1.22e-03	2.44e-02	4.23e-05
rs7633675	3	185510613	IGF2BP2	6.22e-03	1.28e-03	2.44e-02	4.23e-05
rs16860234	3	185510884	IGF2BP2	1.98e-02	6.43e-03	1.00e+00	1.83e-02
rs4402960	3	185511687	IGF2BP2	3.14e-03	6.87e-04	2.05e-02	3.51e-05
rs16860235	3	185512361	IGF2BP2	3.12e-02	1.33e-02	1.00e+00	4.10e-02
rs7640539	3	185513296	IGF2BP2	4.56e-03	1.07e-03	2.44e-02	4.23e-05
rs7651090	3	185513392	IGF2BP2	5.35e-03	9.58e-04	2.19e-02	3.83e-05
rs6444081	3	185514393	IGF2BP2	4.25e-03	1.01e-03	2.44e-02	4.23e-05
rs7646518	3	185514931	IGF2BP2	4.71e-03	1.09e-03	2.44e-02	4.23e-05
rs7637773	3	185515635	IGF2BP2	4.41e-03	1.04e-03	3.00e-02	6.90e-05
rs4686696	3	185516520	IGF2BP2	4.09e-03	9.86e-04	2.44e-02	4.23e-05
rs6767484	3	185520578	IGF2BP2	3.45e-03	9.03e-04	2.44e-02	4.23e-05
rs7640744	3	185522447	IGF2BP2	2.14e-02	8.37e-03	1.00e+00	1.52e-02

rs11711477	3	185526690	IGF2BP2	5.70e-03	1.18e-03	2.74e-02	5.02e-05
rs1470579	3	185529080	IGF2BP2	5.19e-03	9.31e-04	2.98e-02	5.16e-05
rs6769511	3	185530290	IGF2BP2	6.05e-03	1.25e-03	2.99e-02	5.41e-05
rs9859406	3	185534482	IGF2BP2	6.39e-03	1.31e-03	3.16e-02	5.65e-05
rs2548966	5	134215127	TXNDC15	3.01e-02	1.06e-02	1.00e+00	3.97e-02
rs319602	5	134222164	TXNDC15	2.02e-02	7.07e-03	1.00e+00	3.64e-02
rs319598	5	134240235	PCBD2	3.05e-02	1.08e-02	1.00e+00	4.02e-02
rs319592	5	134252619	PCBD2	3.94e-02	1.21e-02	1.00e+00	4.77e-02
rs319589	5	134255333	PCBD2	2.39e-02	9.74e-03	1.00e+00	3.63e-02
rs6883047	5	134272055	PCBD2	2.35e-02	8.55e-03	1.00e+00	4.71e-02
rs7728823	5	134282777	PCBD2	2.43e-02	9.84e-03	1.00e+00	4.02e-02
rs12658264	5	141764189	non-coding	3.84e-02	6.98e-03	1.00e+00	1.35e-01
rs9348440	6	20641336	CDKAL1	3.19e-02	3.36e-03	1.00e+00	1.79e-03
rs6456364	6	20649254	CDKAL1	4.56e-02	4.08e-03	1.00e+00	1.36e-03
rs9295474	6	20652717	CDKAL1	2.56e-03	2.53e-04	6.90e-04	1.85e-07
rs2328545	6	20653550	CDKAL1	2.93e-02	2.50e-03	1.00e+00	7.99e-04
rs9368216	6	20655110	CDKAL1	4.59e-02	4.15e-03	1.00e+00	8.10e-04
rs4712522	6	20656800	CDKAL1	2.32e-03	2.28e-04	6.11e-04	1.79e-07
rs4712523	6	20657564	CDKAL1	2.40e-03	2.36e-04	7.28e-04	3.31e-07
rs4710940	6	20658012	CDKAL1	5.02e-03	8.62e-04	7.01e-03	4.89e-06
rs6906327	6	20659459	CDKAL1	3.76e-03	3.99e-04	6.65e-03	2.12e-06
rs6456367	6	20659587	CDKAL1	1.48e-03	1.31e-04	5.93e-04	1.73e-07
rs6456368	6	20659806	CDKAL1	1.05e-03	5.92e-05	4.31e-04	1.18e-07
rs6456369	6	20660365	CDKAL1	3.60e-03	3.80e-04	6.52e-03	4.56e-06
rs10946398	6	20661034	CDKAL1	2.02e-03	1.13e-04	8.02e-04	2.40e-07
rs7774594	6	20661143	CDKAL1	1.82e-03	1.70e-04	5.91e-04	1.72e-07
rs7754840	6	20661250	CDKAL1	2.09e-03	2.11e-04	1.23e-03	3.70e-07
rs9460544	6	20661529	CDKAL1	2.63e-03	2.60e-04	5.91e-04	1.73e-07
rs9460545	6	20661550	CDKAL1	1.89e-03	1.78e-04	5.91e-04	1.73e-07
rs4712525	6	20662966	CDKAL1	2.48e-03	2.45e-04	6.04e-04	1.74e-07
rs4712526	6	20663035	CDKAL1	1.40e-03	1.22e-04	6.09e-04	1.74e-07
rs9460546	6	20663632	CDKAL1	2.17e-03	2.19e-04	1.23e-03	3.69e-07
rs742642	6	20665081	CDKAL1	2.97e-02	2.55e-03	1.00e+00	7.52e-04
rs7748382	6	20665549	CDKAL1	1.55e-03	1.39e-04	5.91e-04	1.72e-07
rs7772603	6	20665946	CDKAL1	1.76e-03	1.63e-04	5.53e-04	1.58e-07
rs7752780	6	20666022	CDKAL1	1.62e-03	1.47e-04	5.21e-04	1.48e-07
rs7752906	6	20666055	CDKAL1	1.69e-03	1.55e-04	5.13e-04	1.46e-07
rs9358356	6	20667382	CDKAL1	7.29e-04	3.94e-05	4.74e-04	1.32e-07
rs9356743	6	20667688	CDKAL1	1.69e-02	1.99e-03	6.57e-01	2.24e-04
rs9368219	6	20674691	CDKAL1	1.95e-03	4.83e-05	5.23e-05	1.05e-09
rs1012635	6	20675295	CDKAL1	4.27e-02	4.65e-03	8.35e-01	1.98e-03
rs1569699	6	20679310	CDKAL1	4.99e-04	1.55e-05	3.32e-07	1.36e-11
rs7756992	6	20679709	CDKAL1	1.14e-04	4.37e-07	1.05e-08	0.00e+00
rs9350271	6	20683164	CDKAL1	5.42e-04	1.77e-05	7.74e-07	2.58e-11
rs9356744	6	20685486	CDKAL1	5.83e-04	1.98e-05	7.44e-07	2.58e-11
rs7766070	6	20686573	CDKAL1	2.83e-05	1.43e-07	9.06e-09	0.00e+00
rs9368222	6	20686996	CDKAL1	3.81e-05	2.02e-07	9.06e-09	0.00e+00
rs10440833	6	20688121	CDKAL1	1.60e-05	8.06e-08	9.06e-09	0.00e+00
rs2206734	6	20694884	CDKAL1	8.97e-04	3.05e-05	2.44e-03	7.56e-10
rs6931514	6	20703952	CDKAL1	9.05e-05	3.12e-07	9.06e-09	0.00e+00
rs11753081	6	20705590	CDKAL1	6.76e-04	2.20e-05	1.90e-05	4.95e-10
rs1040558	6	20713706	CDKAL1	7.88e-04	2.49e-05	1.66e-05	4.32e-10
rs9295478	6	20716253	CDKAL1	6.95e-03	6.19e-04	1.29e-03	3.00e-07
rs2328548	6	20716958	CDKAL1	6.28e-04	1.32e-05	1.42e-05	3.78e-10
rs6935599	6	20717095	CDKAL1	9.48e-04	3.31e-05	1.42e-05	3.75e-10
rs9465871	6	20717255	CDKAL1	2.63e-04	1.74e-06	6.30e-06	1.25e-10
rs10946403	6	20717404	CDKAL1	8.44e-04	2.78e-05	1.32e-05	3.42e-10
rs2328549	6	20718240	CDKAL1	3.42e-02	2.81e-03	2.02e-01	8.60e-05
rs9358357	6	20719145	CDKAL1	9.98e-04	3.56e-05	1.42e-05	3.73e-10
rs9368224	6	20719232	CDKAL1	3.12e-04	4.08e-06	1.42e-05	3.73e-10
rs9358358	6	20719393	CDKAL1	1.73e-02	2.03e-03	1.23e-01	5.35e-05
rs9460550	6	20719561	CDKAL1	3.59e-04	6.28e-06	1.42e-05	3.73e-10
rs9356746	6	20720279	CDKAL1	1.22e-02	1.84e-03	1.08e-01	4.81e-05
rs9368226	6	20723057	CDKAL1	4.54e-04	1.08e-05	5.48e-05	1.07e-09
rs12111351	6	20724558	CDKAL1	6.75e-03	6.03e-04	3.17e-03	7.70e-07
rs9356747	6	20725007	CDKAL1	7.14e-03	6.36e-04	3.18e-03	7.72e-07
rs9356748	6	20725097	CDKAL1	1.26e-02	1.88e-03	9.40e-02	4.37e-05
rs7767391	6	20725240	CDKAL1	4.07e-04	8.58e-06	6.90e-05	1.39e-09

rs7747752	6	20725423	CDKAL1	5.52e-03	3.43e-04	2.24e-03	5.65e-07
rs9270986	6	32574060	non-coding	3.16e-02	4.82e-03	1.00e+00	2.29e-03
rs9492055	6	129048640	non-coding	4.73e-02	1.51e-02	1.00e+00	4.80e-02
rs11154899	6	137293890	non-coding	4.89e-02	1.11e-02	1.00e+00	1.09e-01
rs10872465	6	137294656	non-coding	4.86e-02	1.10e-02	1.00e+00	1.09e-01
rs2876354	6	137295352	non-coding	4.96e-02	1.13e-02	1.00e+00	1.09e-01
rs11154900	6	137296161	non-coding	4.93e-02	1.12e-02	1.00e+00	1.10e-01
rs6906007	6	137296300	non-coding	3.31e-02	8.27e-03	1.00e+00	8.99e-02
rs10457653	6	137296895	non-coding	4.83e-02	1.09e-02	1.00e+00	1.09e-01
rs10872466	6	137297967	non-coding	3.91e-02	7.35e-03	1.00e+00	9.89e-02
rs4407733	6	137299152	non-coding	4.99e-02	1.14e-02	1.00e+00	1.09e-01
rs947733	6	137304427	non-coding	3.08e-02	8.46e-03	1.00e+00	6.22e-02
rs849133	7	28192280	JAZF1	4.24e-02	1.58e-02	1.00e+00	1.26e-03
rs849134	7	28196222	JAZF1	2.10e-02	7.80e-03	9.84e-01	1.16e-03
rs849135	7	28196413	JAZF1	3.23e-02	1.16e-02	9.75e-01	1.14e-03
rs10281305	7	54890409	non-coding	4.04e-02	1.18e-02	1.00e+00	1.09e-01
rs4493865	7	54898402	non-coding	4.63e-02	1.52e-02	1.00e+00	1.14e-01
rs2442982	8	20590386	non-coding	3.46e-02	2.29e-02	1.00e+00	1.34e-01
rs4734295	8	96000919	non-coding	9.25e-03	3.49e-03	1.00e+00	2.34e-02
rs10113282	8	96038252	C8orf38	3.87e-02	1.02e-02	1.00e+00	4.08e-02
rs1892012	9	19979945	non-coding	4.46e-02	9.54e-03	1.00e+00	6.54e-02
rs10117648	9	19981497	non-coding	4.49e-02	1.03e-02	1.00e+00	6.93e-02
rs7868773	9	19985150	non-coding	4.53e-02	1.04e-02	1.00e+00	7.03e-02
rs10122799	9	19987293	non-coding	4.11e-02	9.64e-03	1.00e+00	9.47e-02
rs10964378	9	19994736	non-coding	4.69e-02	1.37e-02	1.00e+00	1.09e-01
rs10964380	9	19999413	non-coding	4.66e-02	1.72e-02	1.00e+00	1.09e-01
rs7020996	9	22129579	non-coding	1.25e-03	2.02e-04	1.60e-02	3.66e-06
rs2383208	9	22132076	non-coding	2.24e-03	8.61e-05	3.29e-02	3.61e-06
rs10965250	9	22133284	non-coding	1.32e-03	7.13e-05	5.86e-03	7.57e-07
rs10811661	9	22134094	non-coding	1.17e-03	6.51e-05	8.83e-03	7.70e-07
rs1333051	9	22136489	non-coding	2.71e-03	2.70e-04	7.52e-02	3.98e-05
rs2798253	10	94202905	non-coding	1.60e-02	3.19e-03	9.38e-03	9.56e-06
rs6583813	10	94209939	non-coding	3.57e-02	3.13e-03	1.73e-02	1.22e-05
rs11187007	10	94214580	IDE	2.78e-02	2.75e-03	1.26e-02	4.84e-06
rs2149632	10	94232247	IDE	1.94e-02	2.12e-03	1.80e-02	7.72e-06
rs11187033	10	94262359	IDE	1.89e-02	2.07e-03	1.86e-02	7.07e-06
rs10509645	10	94277866	IDE	3.97e-02	3.02e-03	2.55e-02	1.06e-05
rs2421941	10	94345909	non-coding	4.30e-02	1.38e-03	1.00e+00	9.64e-05
rs10786050	10	94367230	KIF11	3.35e-02	7.05e-04	1.00e+00	1.05e-04
rs10882091	10	94374377	KIF11	3.38e-02	7.22e-04	1.00e+00	9.01e-05
rs10882094	10	94387676	KIF11	2.51e-02	6.69e-04	1.00e+00	8.52e-05
rs10882095	10	94394402	KIF11	4.01e-02	3.08e-03	3.16e-02	1.22e-05
rs10736069	10	94395393	KIF11	2.47e-02	6.52e-04	1.00e+00	7.55e-05
rs7900689	10	94395748	KIF11	2.23e-02	4.37e-04	1.00e+00	7.55e-05
rs6583830	10	94398118	KIF11	2.19e-02	4.18e-04	1.00e+00	7.32e-05
rs10882096	10	94401386	KIF11	3.49e-02	3.25e-03	2.74e-02	1.02e-05
rs11187114	10	94406237	KIF11	3.53e-02	3.30e-03	2.44e-02	8.93e-06
rs4933734	10	94414567	KIF11	1.11e-02	2.96e-04	1.00e+00	1.55e-05
rs7911264	10	94436851	non-coding	1.15e-02	3.11e-04	8.34e-02	4.18e-07
rs2488087	10	94446041	non-coding	1.07e-02	1.04e-04	8.34e-02	4.18e-07
rs10882100	10	94460687	non-coding	1.04e-02	9.52e-05	8.65e-02	4.28e-07
rs1111875	10	94462882	non-coding	3.02e-03	3.61e-04	1.24e-03	4.82e-07
rs12778642	10	94464307	non-coding	2.90e-03	2.82e-04	9.05e-04	4.49e-07
rs5015480	10	94465559	non-coding	1.10e-03	7.74e-05	8.78e-04	1.12e-07
rs10882102	10	94466495	non-coding	2.80e-03	3.25e-04	1.30e-03	4.77e-07
rs11187144	10	94469980	non-coding	8.09e-03	1.70e-03	1.03e-02	8.93e-06
rs7087591	10	94473629	non-coding	8.35e-03	1.73e-03	9.41e-03	6.14e-06
rs10748582	10	94477219	non-coding	7.35e-03	1.35e-03	6.65e-03	3.33e-06
rs7923837	10	94481917	non-coding	8.61e-03	1.77e-03	8.97e-03	4.89e-06
rs7923866	10	94482076	non-coding	7.56e-03	1.44e-03	8.37e-03	4.89e-06
rs7917983	10	114732882	TCF7L2	6.55e-03	4.35e-05	1.78e-01	5.65e-07
rs7901275	10	114732906	TCF7L2	4.86e-03	5.33e-05	1.63e-01	3.70e-07
rs4074720	10	114748497	TCF7L2	1.98e-04	9.75e-09	7.42e-09	0.00e+00
rs4074718	10	114748617	TCF7L2	7.27e-05	5.26e-10	6.32e-09	0.00e+00
rs17747324	10	114752503	TCF7L2	6.05e-08	4.00e-13	0.00e+00	0.00e+00
rs7901695	10	114754088	TCF7L2	5.21e-09	8.64e-15	0.00e+00	0.00e+00
rs4506565	10	114756041	TCF7L2	1.07e-10	4.28e-20	0.00e+00	0.00e+00
rs7903146	10	114758349	TCF7L2	2.40e-11	4.61e-22	0.00e+00	0.00e+00

rs10885402	10	114761697	TCF7L2	4.88e-05	3.72e-10	7.42e-09	0.00e+00
rs6585198	10	114762237	TCF7L2	7.86e-05	6.60e-10	7.42e-09	0.00e+00
rs4132670	10	114767771	TCF7L2	3.25e-09	5.32e-15	0.00e+00	0.00e+00
rs6585200	10	114768609	TCF7L2	9.65e-05	9.05e-10	8.10e-09	0.00e+00
rs6585201	10	114768783	TCF7L2	8.37e-05	7.78e-10	8.10e-09	0.00e+00
rs7904519	10	114773927	TCF7L2	1.89e-04	1.68e-09	9.06e-09	0.00e+00
rs10885405	10	114777670	TCF7L2	1.26e-04	1.31e-09	9.06e-09	0.00e+00
rs10885406	10	114777724	TCF7L2	1.32e-04	1.43e-09	9.06e-09	0.00e+00
rs10787472	10	114781297	TCF7L2	1.20e-04	1.18e-09	1.00e-08	0.00e+00
rs7924080	10	114787012	TCF7L2	1.08e-04	1.04e-09	9.06e-09	0.00e+00
rs12243326	10	114788815	TCF7L2	9.89e-09	4.46e-14	0.00e+00	0.00e+00
rs7077039	10	114789077	TCF7L2	1.02e-04	1.19e-10	8.32e-09	0.00e+00
rs7900150	10	114793823	TCF7L2	6.60e-05	8.16e-11	7.42e-09	0.00e+00
rs7100927	10	114796048	TCF7L2	1.79e-04	2.12e-10	7.42e-09	0.00e+00
rs7895340	10	114801525	TCF7L2	5.81e-05	4.77e-11	7.42e-09	0.00e+00
rs11196200	10	114801938	TCF7L2	1.41e-04	1.62e-10	7.42e-09	0.00e+00
rs11196205	10	114807047	TCF7L2	2.08e-04	8.02e-09	7.00e-08	0.00e+00
rs12255372	10	114808902	TCF7L2	2.25e-07	4.11e-12	0.00e+00	0.00e+00
rs12265291	10	114810240	TCF7L2	1.60e-04	4.87e-09	6.65e-08	0.00e+00
rs11196208	10	114811316	TCF7L2	1.51e-04	3.35e-09	6.65e-08	0.00e+00
rs7077247	10	114812071	TCF7L2	1.69e-04	6.27e-09	7.00e-08	0.00e+00
rs12718338	10	114813047	TCF7L2	2.32e-04	1.35e-08	1.11e-07	0.00e+00
rs10832778	11	17394073	B7H6	2.82e-02	1.64e-02	1.00e+00	1.53e-01
rs1557765	11	17403639	non-coding	9.61e-03	6.34e-03	1.00e+00	3.68e-02
rs5215	11	17408630	KCNJ11	8.91e-03	4.50e-03	1.00e+00	2.36e-02
rs7124355	11	17412960	non-coding	2.06e-02	1.57e-02	1.00e+00	4.76e-02
rs757110	11	17418477	ABCC8	9.98e-03	6.16e-03	1.00e+00	2.67e-02
rs1877527	12	71405206	non-coding	4.07e-02	8.08e-03	1.00e+00	3.68e-02
rs11178531	12	71408690	non-coding	2.55e-02	5.63e-03	1.00e+00	2.36e-02
rs7132840	12	71411561	non-coding	3.81e-02	9.35e-03	1.00e+00	2.19e-02
rs2063591	12	71411855	non-coding	2.59e-02	5.72e-03	1.00e+00	3.59e-02
rs7957932	12	71421552	non-coding	4.76e-02	1.31e-02	1.00e+00	4.31e-02
rs1512991	12	71422768	non-coding	1.77e-02	4.99e-03	1.00e+00	2.33e-02
rs7956274	12	71424402	non-coding	1.81e-02	5.08e-03	1.00e+00	2.36e-02
rs7959965	12	71425164	non-coding	1.85e-02	5.17e-03	1.00e+00	2.36e-02
rs7298255	12	71428069	non-coding	1.30e-02	4.01e-03	1.00e+00	2.07e-02
rs10784891	12	71429798	non-coding	1.64e-02	4.73e-03	1.00e+00	2.79e-02
rs7955901	12	71433293	non-coding	2.75e-02	6.07e-03	1.00e+00	3.17e-02
rs4760894	12	71438923	non-coding	2.90e-02	7.53e-03	1.00e+00	3.64e-02
rs4760785	12	71438945	non-coding	2.63e-02	5.81e-03	1.00e+00	3.69e-02
rs4760895	12	71439127	non-coding	2.67e-02	5.90e-03	1.00e+00	3.77e-02
rs7138300	12	71439589	non-coding	2.71e-02	5.99e-03	1.00e+00	3.86e-02
rs1913201	12	71439825	non-coding	2.86e-02	7.44e-03	1.00e+00	3.86e-02
rs10879240	12	71443285	non-coding	3.27e-02	8.85e-03	1.00e+00	3.88e-02
rs7313973	12	71444058	non-coding	1.56e-02	4.58e-03	1.00e+00	4.02e-02
rs1554522	17	25913172	KSR1	4.36e-02	1.45e-02	1.00e+00	2.13e-01

S12 Details of simulations results in Section 6.1

Figure S4 shows the false discovery proportion (FDP) in a replicability analysis (top), and in an analysis to discover associations (bottom). The variation in FDP decreases with M, and is very small for $M = 100,000$.
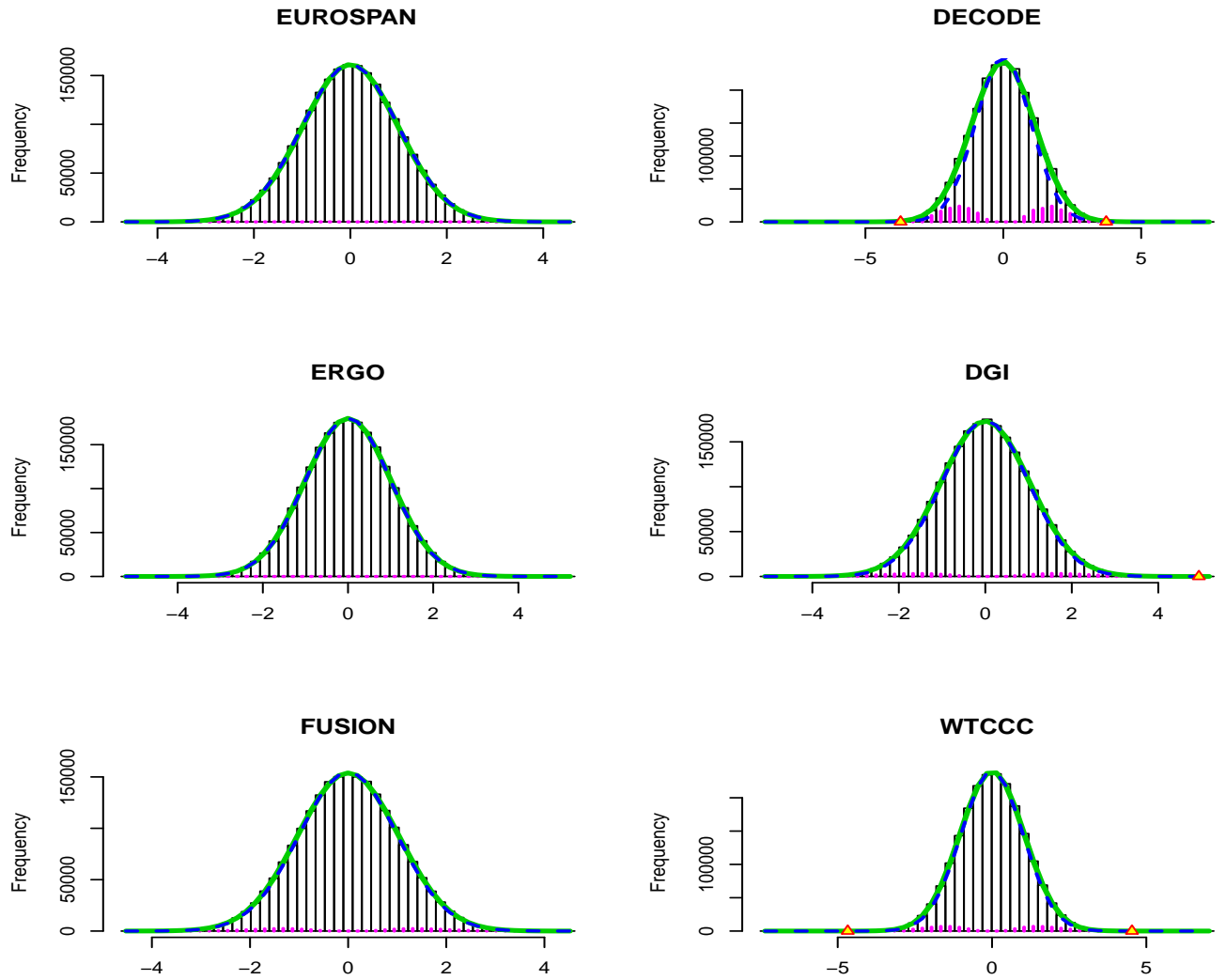
Figure S3: The histogram of z-scores for each of the six T2D GWA studies. The heavy curve is the estimate $\hat{f}_i(z)$ for the mixture density $f(z)$, scaled to match the histogram area. Dashed curve is scaled estimate $\hat{\pi}_0(i)f_0(z)$, where $f_0(z)$ is the standard normal density. The estimated non-null counts are shown in pink.
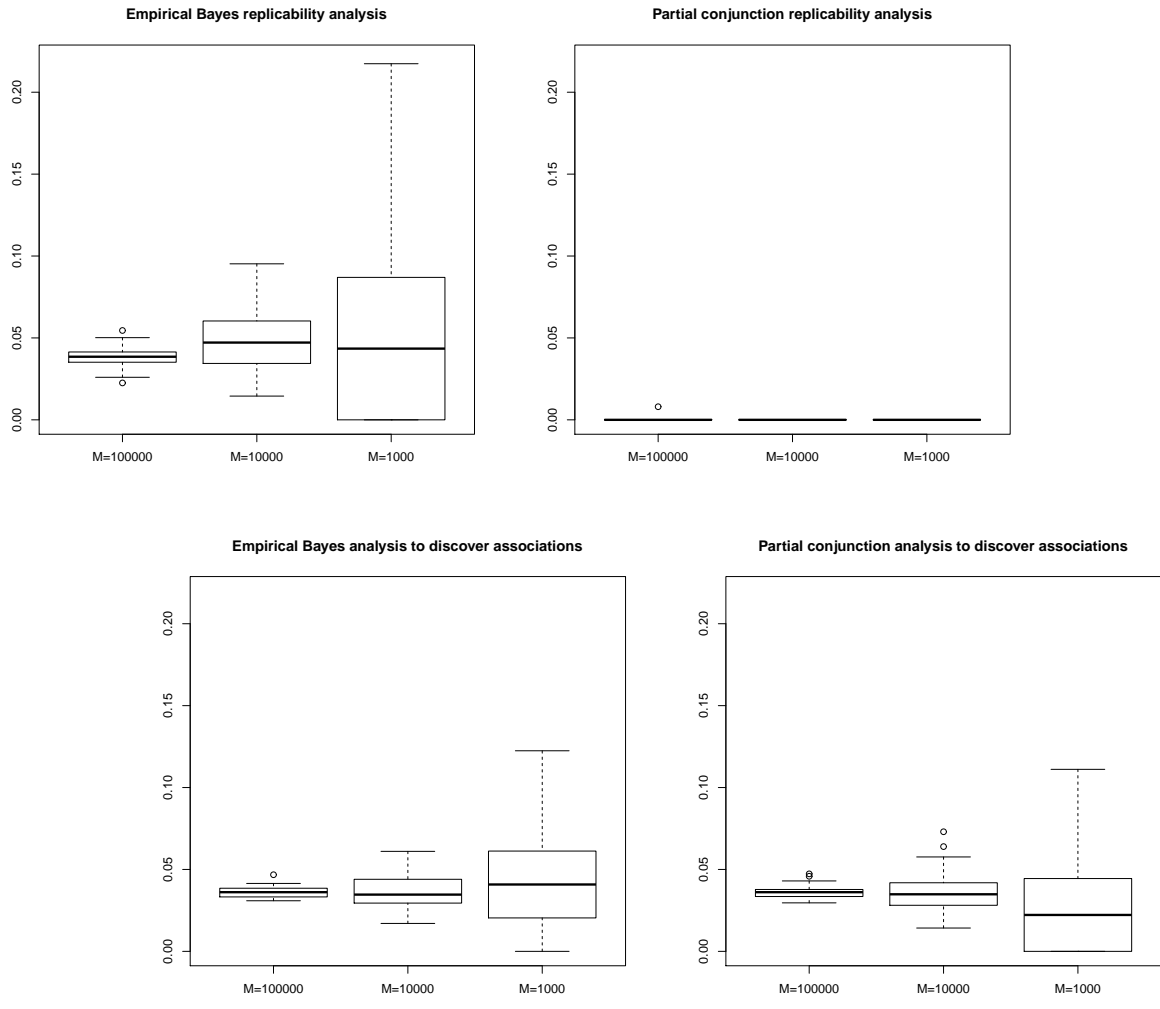
Figure S4: Replicability analysis (top) and analysis to discover association (bottom): Boxplots of FDP for M=100000, M=10000, and M=1000 for empirical Bayes analysis (left) and the analysis of BHY09 (right).